Theses and Dissertations 1. Thesis and Dissertation Collection, all items

1987-09

Tracking procedure for non-normally distributed measurement errors

Kukliansky, Alexander

http://hdl.handle.net/10945/22558

http://www.nps.edu/library

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

TRACKING PROCEDURE
FOR
NON-NORMALLY DISTRIBUTED
MEASUREMENT ERRORS

by

Alexander Kukliansky

September 1987

Thesis Advisor   Donald P. Gaver

Approved for public release; distribution is unlimited.

T234278

REPORT DOCUMENTATION PAGE

REPORT SECURITY CLASSIFICATION	1b RESTRICTIVE MARKINGS
UNCLASSIFIED	
SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT
	Approved for public release;
DECLASSIFICATION / DOWNGRADING SCHEDULE	distribution is unlimited

PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)

NAME OF PERFORMING ORGANIZATION	6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION
val Postgraduate School	55	Naval Postgraduate School

ADDRESS (City State and ZIP Code)	7b ADDRESS (City, State and ZIP Code)
nterey, California 93943-5000	Monterey, California 93943-5000

NAME OF FUNDING / SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

ADDRESS (City State and ZIP Code)	10 SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

TITLE (include Security Classification)

ACKING PROCEDURE FOR NON-NORMALLY DISTRIBUTED MEASUREMENT ERRORS

PERSONAL AUTHOR(S)

KUKLIANSKY, Alexander

TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year Month Day)	15 PAGE COUNT
ter's Thesis	FROM ___ TO ___	1987 September	59

SUPPLEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Tracking, Kalman Filter, Robust Procedure, Outliers, non-normal distributions.

ABSTRACT (Continue on reverse if necessary and identify by block number)

The Kalman Filter is a widely used procedure in tracking gorithms. When normality assumptions are violated, the Kalman Filter rformance tends to degrade. In this thesis a new procedure is troduced for accommodating non-normal properties of measurement error stributions. The procedure is developed for the multi-observer tuation. Simulation experiment results are presented and numerical mparisons are made between the Kalman Filter performance and that of e new procedure.

DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED
NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c OFFICE SYMBOL
f. Donald P. Gaver	408-646-2605	55Gv

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

Tracking procedure
for
non-normally distributed measurement errors

by

Alexander Kukliansky
Lieutenant Commander, Israeli Navy
B.Sc., Hebrew University of Jerusalem, 1976

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1987

ABSTRACT

The Kalman Filter is a widely used procedure in tracking algorithms. When normality assumptions are violated, the Kalman Filter performance tends to degrade. In this thesis a new procedure is introduced for accommodating non-normal properties of measurement error distributions. The procedure is developed for the multi-observer situation. Simulation experiment results are presented and numerical comparisons are made between the Kalman Filter performance and that of the new procedure.

TABLE OF CONTENTS

4

LIST OF TABLES

LIST OF FIGURES

DEDICATION

This work is dedicated to my wife. Her patience and devotion made this work possible.

ACKNOWLEDGMENT

I. INTRODUCTION

Tracking algorithms are widely used in modern technology. Perhaps the most commonly used algorithm is the Kalman Filter (KF) (cf. [Ref. 1] and [Ref. 2]) which is based on the following assumptions:

a) The target movement model is known, and has normally distributed fluctuations.

b) The sensors tracking the target provide unbiased, normally distributed measurements of current target "position".

If the assumptions are satisfied, KF is the best algorithm possible. In practice, however, the normality assumption may not be justified, and the KF performance is degraded. A number of modifications of the KF have been made by different authors based either on heuristics, analytic derivations, or both in order to improve KF performance in different situations (cf. [Ref. 3] and [Ref. 4]).

In this thesis a tracking procedure is developed for the following one-dimentional basic (and classical) model:

a) $\quad \theta_{n+1} = \theta_n + \varepsilon_{n+1}; \qquad \varepsilon_{n+1} \sim N(0,\tau^2); \qquad n = 0,1,2,...\infty$

b) $\quad y_{n+1,j} = \theta_{n+1} + \delta_{n+1,j}; \quad \delta_{n+1,j} \sim t(d_j,\sigma_j); \qquad j = 1,2,3,....m \quad n = 0,1,2,....\infty$

where

- θ_{n+1} is the target position after time step n
- $y_{n+1,j}$ is the measurement of target "position" obtained by sensor j at time step $n+1$.

In the model the target performs a random walk with normally distributed steps with mean zero and known variance τ^2, and the jth sensor's measurement errors are unbiased and t-distributed with known parameters (d_j,σ_j) where d_j is the degrees of freedom and σ_j is the scale parameter of the t-distribution. There are m sensors.

This model violates the classical KF assumptions by allowing thick-tailed (outlier-prone) measurement errors. This model is designed for the situation where m is small, and it isn't possible to make use of the central limit theorem effect.

The mathematical tractability of the model is complicated by the fact that there is no simple analytic form for an estimate of location of the target with Student-t measurement errors.

10

In Chapter II two possible ways to derive the classical KF equations are given. These derivations will help to motivate the procedure suggested in Chapter III. Some practical considerations regarding the implementation of the procedure will be given in Chapter III as well.

In Chapter IV the simulation experiment implementing the procedure derived in Chapter III is described and numerical performance comparisons of the KF and the new procedure are presented.

Chapter V contains final remarks about the new filter and further topics for future development.

Appendix A contains tabulated results of simulation experiment.

Appendix B contains details on simulation experiment implementation.

Appendix C contains graphical presentation of the results from Appendix A.

II. KALMAN FILTER

The basic model for the classical Kalman filter is that the target moves according to a random walk with normally distributed step sizes. The measurement errors are also normal and statistically independent from target "position", i.e.:

a) $\theta_{n+1} = \theta_n + \varepsilon_{n+1}$; $\qquad \varepsilon_{n+1} \sim N(0, \tau^2)$; $\quad n = 0,1,2,...\infty$

b) $y_{n+1,j} = \theta_{n+1} + \delta_{n+1,j}$; $\quad \delta_{n+1,j} \sim N(0, v_j)$; $\quad j = 1,2,3,...m \quad n = 0,1,2,...\infty$

There are three alternative ways to obtain the KF equations:

1. the maximum likelihood estimation approach.
2. the minimum variance unbiased estimation approach.
3. the minimization of the mean square of estimation error $(\theta_n - \theta_n)$.

Extension and synergism of the first two methods will lead us to the procedure described in Chapter III.

A. MAXIMUM LIKELIHOOD

Let $\hat{\theta}_n$ be the estimate of θ_n after time step n and C_n be its variance. Since $\hat{\theta}_n$ is normally[1] distributed with mean θ_n and variance C_n, the conditional likelihood of θ_{n+1} given $\hat{\theta}_n$ and $y_{n+1,1},...,y_{n+1,m}$ is:

$$L(\theta_{n+1}|\hat{\theta}_n, y_{n+1,1},...,y_{n+1,m}) = e^{-\frac{1}{2}\frac{(\theta_{n+1}-\hat{\theta}_n)^2}{v_0}} * \prod_{j=1}^{m} e^{-\frac{1}{2}\frac{(y_{n+1,j}-\theta_{n+1})^2}{v_j}} \qquad (2.1)$$

where $v_0 = v_0(n) = C_n + \tau^2$
and $v_j = \sigma_j^2 = \text{Var}(y_j)$ $\quad (j = 1,...,m)$.

---

[1]In the KF case normality is preserved at each estimation step as can be easily verified.

Finding the maximum for this likelihood is equivalent to finding minimum of the exponent.

i.e.

$$min \left(\sum_{j=0}^{m} \frac{(\theta_{n+1} - y_j)^2}{v_j} \right) = Q(\theta_{n+1}) \qquad (2.2)$$

(dropping the first subscript on the y_{n+1}'s and using $y_0 = \hat{\theta}_n$ for a compact notation). Expanding the squares we obtain

$$Q(\theta_{n+1}) = \theta_{n+1}^2 \sum_{j=0}^{m} \frac{1}{v_j} - 2\theta_{n+1} \sum_{j=0}^{m} \frac{y_j}{v_j} + \sum_{j=0}^{m} \frac{y_j^2}{v_j} \qquad (2.3)$$

This simple quadratic expression is minimized by letting

$$(K.1) \qquad \hat{\theta}_{n+1} = \frac{\displaystyle\sum_{j=0}^{m} \frac{y_j}{v_j}}{\displaystyle\sum_{j=0}^{m} \frac{1}{v_j}} = \frac{\displaystyle\frac{\hat{\theta}_n}{C_n + \tau^2} + \sum_{j=1}^{m} \frac{y_{n+1,j}}{v_j}}{\displaystyle\frac{1}{C_n + \tau^2} + \sum_{j=1}^{m} \frac{1}{v_j}} .$$

Since $(K.1)$ is a linear equation we may immediately obtain the variance of the estimate.

$$(K.2) \qquad C_{n+1} = Var \hat{\theta}_{n+1} = \frac{Var \left(\displaystyle\sum_{j=0}^{m} \frac{y_j}{v_j} \right)}{\left(\displaystyle\sum_{j=0}^{m} \frac{1}{v_j} \right)^2} = \frac{\displaystyle\sum_{j=0}^{m} \frac{v_j}{v_j^2}}{\left(\displaystyle\sum_{j=0}^{m} \frac{1}{v_j} \right)^2} = \frac{1}{\displaystyle\sum_{j=0}^{m} \frac{1}{v_j}}$$

Equations $(K.1)$ and $(K.2)$ are the only ones needed to implement the KF.

B.   MINIMUM VARIANCE

Given $y_0 = \hat{\theta}_n$ and $y_j = y_{n+1,j}$ $j = 1,....,m$; it is desired to get an unbiased minimum variance estimate (UMVE) of θ_{n+1} .

Since $y_0, y_1, ... y_m$ are all unbiased estimates of θ_{n+1}, any linear combination

$$H = a_0 y_0 + a_1 y_1 + ... + a_m y_m \qquad (2.4)$$

constrained by $\sum a_j = 1$ will also be an unbiased estimate of θ_{n+1} .

13

Since

$$Var(\Pi) = a_0^2 v_0 + a_1^2 v_1 + \ldots + a_m^2 v_m \quad, \tag{2.5}$$

finding the UMVE is equivalent to solving the following problem:

$$min \; V = a_0^2 v_0 + a_1^2 v_1 + \ldots + a_m^2 v_m$$
$$S.T. \; a_0 + a_1 + \ldots + a_m = 1 \tag{2.6}$$

Given a Lagrange multiplier λ, the Lagrangian equation may be expressed as:

$$\mathcal{L} = a_0^2 v_0 + a_1^2 v_1 + \ldots + a_m^2 v_m - \lambda(a_0 + a_1 + \ldots + a_n - 1). \tag{2.7}$$

After taking partial derivatives and setting them equal to zero the resulting system of $m+2$ linear equations is:

$$2a_0 v_0 - \lambda = 0$$
$$2a_1 v_1 - \lambda = 0$$
$$\ldots$$
$$\ldots$$
$$2a_m v_m - \lambda = 0 \tag{2.8}$$
$$-a_0 - a_1 - \ldots - a_m = -1.$$

Multiplying each equation by the corresponding $1/(2v_j)$ and adding all the equations together results in:

$$\lambda = \frac{2}{\sum\limits_{j=0}^{m} \frac{1}{v_j}} \tag{2.9}$$

Substitution then leads to

$$a_j = \frac{\frac{1}{v_j}}{\sum\limits_{j=0}^{m} \frac{1}{v_j}} \tag{2.10}$$

14

So, once again the final equations are equations $(K.1)$ and $(K.2)$

$(K.1)$
$$\hat{\theta}_{n+1} = H = \frac{\displaystyle\sum_{j=0}^{m} \frac{y_j}{v_j}}{\displaystyle\sum_{j=0}^{m} \frac{1}{v_j}}$$

$(K.2)$
$$C_{n+1} = min\ V = \frac{1}{\displaystyle\sum_{j=0}^{m} \frac{1}{v_j}}$$

Note that the minimum variance approach does not require normality assumptions, but normality assumptions make it equivalent to the maximum likelihood approach.

15

III. W-FILTER

Now to return to the basic model, target motion is a random walk with normal step size, but the m sensors have independent Student-t distributed measurement errors.

The basic model:

a) $\quad \theta_{n+1} = \theta_n + \varepsilon_{n+1}; \qquad \varepsilon_{n+1} \sim N(0, \tau^2); \quad n = 0, 1, 2, \dots \infty$

b) $\quad y_{n+1,j} = \theta_{n+1} + \delta_{n+1,j}; \quad \delta_{n+1,j} \sim t(d_j, \sigma_j); \quad j = 1, 2, 3, \dots m \quad n = 0, 1, 2, \dots \infty$

It is assumed that our procedure is producing unbiased approximately normally distributed estimates. This means that after time step n, $\hat{\theta}_n$ is approximately normally distributed with mean θ_n and variance C_n. This is the same assumption as used by [Ref. 3] and [Ref. 4]. Based on this assumption we need to construct $\hat{\theta}_{n+1}$ and C_{n+1}.

A. THE BEST $\hat{\theta}_{n+1}$

1. The Criterion

Based on the assumption above, the conditional likelihood of θ_{n+1} given $\hat{\theta}_n$ and $y_{n+1,1}, \dots, y_{n+1,m}$ has the form:

$$L(\theta_{n+1} | \hat{\theta}_n, y_{n+1,1}, \dots, y_{n+1,m}) = e^{-\frac{1}{2} \frac{(\theta_{n+1} - \hat{\theta}_n)^2}{C_n + \tau^2}} \times \prod_{j=1}^{m} \frac{c(d_j)}{\left[1 + \left(\frac{y_{n+1,j} - \theta_{n+1}}{\sigma_j} \right)^2 \frac{1}{d_j} \right]^{\frac{d_j + 1}{2}}} \qquad (3.1)$$

where $c(d_j)$ is some constant depending on the degrees-of-freedom parameter d_j of the jth sensor. Although it is possible to look for the maximum likelihood point, taking that approach has the following two interrelated potential problems. First there may be multiple local maxima present. In an extreme case there may be as many as $m+1$ local maxima. (See [Ref. 4] for local maxima conditions analysis when $m = 1$.) Second, in the multiple maxima situation, the global maxima may be tall and skinny, not having much likelihood to support it. One would not like to commit all belief on the basis of such evidence.

To deal with these problems use the following (smoothing) approach will be used: instead of choosing $\hat{\theta}_{n+1}$ to maximize the probability of having θ_{n+1} in an infinitesimal probability element $d\theta$ around $\hat{\theta}_{n+1}$, the estimator $\hat{\theta}_{n+1}$ will be chosen so as to maximize the probability of having θ_{n+1} in an interval $[\hat{\theta}_{n+1} - w, \hat{\theta}_{n+1} + w]$, the later interval having a finite (usually non-zero) predetermined length $2w$. In other words the concern is not having an estimate exactly "on the spot" but rather having it within distance w from true value of θ_{n+1}.

In mathematical terms the approach will be to find the solution to the equation:

$$(W.1.a) \qquad \max_{\theta_{n+1}} I(\theta_{n+1}) = \max_{\theta_{n+1}} \int_{\theta_{n+1}-w}^{\theta_{n+1}+w} L(\zeta | \hat{\theta}_n, y_{n+1,1},...,y_{n+1,m}) \, d\zeta$$

In the equation $(W.1.a)$ w is serving as a tuning parameter and has the following effect on the solution:

- When $w = 0$ solving the equation $(W.1.a)$ is equivalent to finding maximum likelihood point.

- When $w > 0$ and finite, equation $(W.1.a)$ tends to down-weight skinny peaks more than fat ones.

- When w is large enough, equation $(W.1.a)$ will have a unique[2] solution. However it is generally neither required nor optimal to use very large values for w.

- When w approaches infinity, any value of $\hat{\theta}_{n+1}$ will satisfy the equation.

The idea of using a non-zero w has additional appeal since, in some practical situations, occurrence of a large tracking error may degrade sensor performance for consecutive measurements. (Such sensor dependence on tracking performance is not reflected in the model.)

2. The Solution Technique

Finding a solution for equation $(W.1.a)$ directly involves an exhaustive search with numerical integration; not a very exciting prospect at first sight. However it has the advantage of algorithmic simplicity and guaranteed global maxima.

---

[2] Since the likelihood function is positive and asymptotically approaches zero when θ_{n+1} approaches $\pm \infty$ this statement can be easily proved.

a. Analytic/iterative approach

Alternatively, we may try to find an analytic solution to the equation ($IV.1.a$) by taking the derivative of $I(\theta_{n+1})$ with respect to θ_{n+1} and setting it equal to zero.

$$\frac{dI(\theta_{n+1})}{d\theta_{n+1}} = e^{-\frac{1}{2}\frac{(\theta_{n+1}+w-\hat{\theta}_n)^2}{C_n+\tau^2}} * \prod_{j=1}^{m} \frac{c(d_j)}{\left[1+\left(\frac{y_{n+1,j}-w-\theta_{n+1}}{\sigma_j}\right)^2 \frac{1}{d_j}\right]^{\frac{d_j-1}{2}}} -$$

$$e^{-\frac{1}{2}\frac{(\theta_{n+1}-w-\hat{\theta}_n)^2}{C_n+\tau^2}} * \prod_{j=1}^{m} \frac{c(d_j)}{\left[1+\left(\frac{y_{n+1,j}+w-\theta_{n+1}}{\sigma_j}\right)^2 \frac{1}{d_j}\right]^{\frac{d_j+1}{2}}}$$

(3.2)

After setting the derivative equal to zero and performing cancellations

$$e^{-\frac{1}{2}\frac{(\hat{\theta}_{n+1}+w-\hat{\theta}_n)^2}{C_n+\tau^2}} * \prod_{j=1}^{m} \frac{1}{\left[1+\left(\frac{y_{n+1,j}-w-\hat{\theta}_{n+1}}{\sigma_j}\right)^2 \frac{1}{d_j}\right]^{\frac{d_j+1}{2}}} =$$

(3.3)

$$e^{-\frac{1}{2}\frac{(\hat{\theta}_{n+1}-w-\hat{\theta}_n)^2}{C_n+\tau^2}} * \prod_{j=1}^{m} \frac{1}{\left[1+\left(\frac{y_{n+1,j}+w-\hat{\theta}_{n+1}}{\sigma_j}\right)^2 \frac{1}{d_j}\right]^{\frac{d_j+1}{2}}}$$

18

Next, taking logs of both sides of this equation results in:

$$
-\frac{1}{2}\frac{(\hat{\theta}_{n+1}+w-\hat{\theta}_n)^2}{C_n+\tau^2} - \sum_{j=1}^{m}\frac{d_j+1}{2}\ln\left[1+\left(\frac{y_{n+1,j}-w-\hat{\theta}_{n+1}}{\sigma_j}\right)^2\frac{1}{d_j}\right] =
$$

$$
-\frac{1}{2}\frac{(\hat{\theta}_{n+1}-w-\hat{\theta}_n)^2}{C_n+\tau^2} - \sum_{j=1}^{m}\frac{d_j+1}{2}\ln\left[1+\left(\frac{y_{n+1,j}+w-\hat{\theta}_{n+1}}{\sigma_j}\right)^2\frac{1}{d_j}\right]
$$

(3.4)

Multiplying through by (-1), expanding the squares, and pulling out the common denominator of the logarithm gives:

$$
\frac{\hat{\theta}_{n+1}^2+w^2+\hat{\theta}_n^2+2\hat{\theta}_{n+1}w-2\hat{\theta}_nw-2\hat{\theta}_{n+1}\hat{\theta}_n}{C_n+\tau^2} + \sum_{j=1}^{m}(d_j+1)\ln[d_j\sigma_j^2+(y_{n+1,j}-w-\hat{\theta}_{n+1})^2] - \sum_{j=1}^{m}\ln d_j\sigma_j^2 =
$$

(3.5)

$$
\frac{\hat{\theta}_{n+1}^2+w^2+\hat{\theta}_n^2-2\hat{\theta}_{n+1}w+2\hat{\theta}_nw-2\hat{\theta}_{n+1}\hat{\theta}_n}{C_n+\tau^2} + \sum_{j=1}^{m}(d_j+1)\ln[d_j\sigma_j^2+(y_{n+1,j}+w-\hat{\theta}_{n+1})^2] - \sum_{j=1}^{m}\ln d_j\sigma_j^2
$$

Collecting terms and rearranging leads to

$$
\frac{4\hat{\theta}_{n+1}w}{C_n+\tau^2} = \frac{4\theta_n w}{C_n+\tau^2} - \sum_{j=1}^{m}(d_j+1)\ln\frac{d_j\sigma_j^2+(y_{n+1,j}-w-\hat{\theta}_{n+1})^2}{d_j\sigma_j^2+(y_{n+1,j}-w-\hat{\theta}_{n+1})^2}
$$

(3.6)

And finally,

$$
(W\,1.b)\qquad \hat{\theta}_{n+1} = \hat{\theta}_n + \frac{C_n+\tau^2}{4w}\sum_{j=1}^{m}(d_j-1)\ln\frac{d_j\sigma_j^2+(y_{n+1,j}+w-\hat{\theta}_{n+1})^2}{d_j\sigma_j^2+(y_{n+1,j}-w-\hat{\theta}_{n+1})^2}
$$

Equation $(W.1.b)$ is the first order condition for a local maximum for equation $(W.1.a)$. But perhaps a more natural way to check the first order conditions at point $\hat{\theta}_{n+1}$ is to see if the likelihood function takes equal values at points $\hat{\theta}_{n+1}-w$ and $\hat{\theta}_{n+1}+w$.

19

When w approaches zero, equation ($IV.1.b$) approaches the form:

$$\hat{\theta}_{n+1} = \hat{\theta}_n + (C_n + \tau^2) \sum_{j=1}^{m} \frac{(d_j+1)(y_{n+1,j} - \hat{\theta}_{n+1})}{d_j \sigma_j^2 + (y_{n+1,j} - \hat{\theta}_{n+1})^2} \tag{3.7}$$

(The right-most term is the first derivative with respect to $\hat{\theta}_{n+1}$ of the logarithmic function.)

This is exactly the first-order condition for a maximum likelihood parameter value.

Since $\hat{\theta}_{n+1}$ appears in the logarithmic term of ($IV.1.b$), equation ($IV.1.b$) is difficult to solve directly. One solution procedure is to use ($IV.1.b$) in a recursive manner to obtain $\hat{\theta}_{n+1}$. The empirical experience with this approach indicates that unless w is relatively large, direct application of equation ($IV.1.b$) does not converge[3] in most cases.

b. Integral evaluation approach

Instead of using a sophisticated iterative procedure, the simple exhaustive search procedure described below will be used.

Define

$$\theta_{min} = min(\hat{\theta}_n, y_{n+1,1}, \cdots, y_{n+1,m})$$
$$\theta_{max} = max(\hat{\theta}_n, y_{n+1,1}, \cdots, y_{n+1,m}) \tag{3.8}$$

The interval $T = [\theta_{min}, \theta_{max}]$ includes all reasonable candidate points for $\hat{\theta}_{n+1}$. It is possible to evaluate equation ($IV.1.b$) at a finite number of points (say 20) from the interval T in accordance with the resolution required, and take the one with smallest violation of equation ($IV.1.b$) as the estimate $\hat{\theta}_{n+1}$.

The approach of picking the point with the minimum violation of equation ($IV.1.b$) does not guarantee even local maxima. However in some of the simulations that were performed it proved satisfactory unless there were many cases when the likelihood function had large "flat spots" and w was small. This generally happens when the measurement variances are large and/or the number of observers m is large.

A procedure based on evaluation of first or higher order derivatives will be computationally costly and in general will be more difficult to solve because of local maxima and "flat spots" in the likelihood function.

---

[3]The "jumps" between iterations may be very large and the estimate may reach totally unreasonable values.

Therefore, the safest procedure for finding the solution of equation ($IV.1.a$) is exhaustive search over the interval T with numerical integral evaluation. This is the procedure used in the simulation studies reported in the next Chapter. In the simulation, search and integral evaluation were performed in a single DO-LOOP (See Appendix B for details on implementation). The simulation was not written with computational efficiency in mind, nevertheless the computational burden did not reach an unacceptable level.

B. VARIANCE OF $\hat{\theta}_{n+1}$

In order to keep the filter going once $\hat{\theta}_{n+1}$ has been obtained, it is necessary to compute its variance C_{n+1}. Since equation ($IV.1.a$) is implicit and equation ($IV.1.b$) is not linear, there is no natural analogy to equation ($K.2$).

1. Linearization approach

One possible approach to calculate the variance of $\hat{\theta}_{n+1}$ is to approximate equation ($IV.1.b$) with a linear equation obtained by a first order Taylor expansion of the logarithmic terms around $\hat{\theta}_{n+1} = y_{n+1,j}$ for each j. Cancelling and rearranging terms results in the following approximation to equation ($IV.1.b$):

($W.1.c$)
$$\hat{\theta}_{n+1} = K^{-1} (\hat{\theta}_n + (C_n + \tau^2) \sum_{j=1}^{m} k_j y_{n+1,j})$$

where

$$k_j = \frac{d_j + 1}{d_j \sigma_j^2 + w^2} \tag{3.9}$$

$$K = 1 + (C_n + \tau^2) \sum_{j=1}^{m} k_j \tag{3.10}$$

This leads to:

$$C_{n+1} = K^{-2} (C_n + (C_n + \tau^2)^2 \sum_{j=1}^{m} k_j^2 v_j) \tag{3.11}$$

where $v_j = \text{Var}(y_{n+1,j})$.

Empirical evidence indicates that in practice this approach performs very badly, giving values for C_n which are far too small. The C_n depends only on n and converges to a value for large n.

2. Minimization approach

a. *The Idea*

Having rejected the above linearization approach, a different one is introduced by the following motivation.

The "best" estimate, $\hat{\theta}_{n+1}$, is some combination of $\hat{\theta}_n, y_{n+1,1}, \dots, y_{n+1,m}$. In the case of the KF it is possible to pick a linear combination having minimal variance and remain consistent with maximum likelihood criteria. In the WF case this consistency no longer holds, since generally the estimate obtained by finding the solution to the equation ($IV.1.a$) will differ from the minimum variance estimate. But in this case it is still possible find a minimum variance linear combination constrained to yield the "best" (already known) value for $\hat{\theta}_{n+1}$.

This means that in order to obtain an estimate of variance C_{n+1} we have to solve the following constrained minimization problem:

The symbolic formulation would be

$$min \ V = a_0^2 v_0 + a_1^2 v_1 + \dots + a_m^2 v_m \tag{3.12}$$

$$S.T. \ \ a_0\hat{\theta}_n + a_1 y_{n+1,1} + \dots + a_m y_{n+1,m} = \hat{\theta}_{n+1}$$

$$a_0 + a_1 + \dots + a_m = 1$$

where

$$v_0 = C_n + \tau^2 \tag{3.13}$$

$$v_j = Var(y_{n+1,j}) = \frac{d_j}{d_j - 2} \sigma_j^2 \ ; \qquad j = 1,2,\dots,m \tag{3.14}$$

b. *Formula derivation*

If $m = 1$, then no minimization is necessary and the problem has a unique solution.

For notational simplicity the first subscript will be dropped, and let $y_0 = \hat{\theta}_n$

and $\hat{\theta} = \hat{\theta}_{n+1}$. Let λ_1 and λ_2 be Lagrange multipliers. The Lagrangian formulation is then:

$$\mathscr{L} = \sum_{j=0}^{m} a_j^2 v_j \ - \lambda_1 \left(\sum_{j=0}^{m} a_j y_j - \hat{\theta} \right) - \lambda_2 \left(\sum_{j=0}^{m} a_j - 1 \right) \qquad (3.15)$$

Taking partial derivatives with respect to a_j and λ_1, λ_2 to obtain a system of $m+3$ linear equations with $m+3$ unknowns gives:

$$2v_0 a_0 \ - \lambda_1 y_0 \ - \lambda_2 \ = 0$$
$$2v_1 a_1 \ - \lambda_1 y_1 \ - \lambda_2 \ = 0$$
$$\vdots \quad \vdots \quad \vdots$$
$$\vdots \quad \vdots \quad \vdots$$
$$2v_m a_m \ - \lambda_1 y_m \ - \lambda_2 \ = 0 \qquad (3.16)$$

$$\sum_{j=0}^{m} a_j y_j = \hat{\theta}$$

$$\sum_{j=0}^{m} a_j = 1$$

Multiplying each one of the first $m+1$ equations by the corresponding $1/v_j$ and adding all together results in:

$$-\lambda_1 \sum_{j=0}^{m} \frac{y_j}{v_j} - \lambda_2 \sum_{j=0}^{m} \frac{1}{v_j} = \ -2 \qquad (3.17)$$

Since $\sum a_j = 1$.

Multiplying each one of the first $m+1$ equations by corresponding y_j/v_j, the $(m+2)^{nd}$ equation by (-2) and adding all together produces:

$$-\lambda_1 \sum_{j=0}^{m} \frac{y_j^2}{v_j} - \lambda_2 \sum_{j=0}^{m} \frac{y_j}{v_j} = \ -2\hat{\theta} \qquad (3.18)$$

23

After changing signs a system of two linear equations follows:

$$\lambda_1 A + \lambda_2 B = 2\hat{\theta}$$
$$\lambda_1 B + \lambda_2 C = 2 \qquad\qquad (3.19)$$

where

$$A = \sum_{j=0}^{m} \frac{y_j^2}{v_j} \quad ; \quad B = \sum_{j=0}^{m} \frac{y_j}{v_j} \quad ; \quad C = \sum_{j=0}^{m} \frac{1}{v_j}$$

Applying Cramer's rule gives

$$\lambda_1 = \frac{2\hat{\theta}C - 2B}{AC - B^2} \qquad\qquad (3.20)$$

$$\lambda_2 = \frac{2A - 2\hat{\theta}B}{AC - B^2} \qquad\qquad (3.21)$$

Substitution yields

$$a_j = \frac{\lambda_1 y_j + \lambda_2}{2v_j} \qquad (j=0,1,2,...,m) \qquad (3.22)$$

or explicitly

$$a_j = \frac{y_j\left(\hat{\theta}\sum_{j=0}^{m}\frac{1}{v_j} - \sum_{j=0}^{m}\frac{y_j}{v_j}\right) + \left(\sum_{j=0}^{m}\frac{y_j^2}{v_j} - \hat{\theta}\sum_{j=0}^{m}\frac{y_j}{v_j}\right)}{v_j\left(\sum_{j=0}^{m}\frac{y_j^2}{v_j}\sum_{j=0}^{m}\frac{1}{v_j} - \left(\sum_{j=0}^{m}\frac{y_j}{v_j}\right)^2\right)} \qquad (3.23)$$

Note that the denominator is equal to zero in case $y_0 = y_1 = ... = y_m$, which is a highly improbable event.

Since the Hessian matrix

$$
\begin{matrix}
2v_0 & 0 & \dots & 0 \\
0 & 2v_1 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & \dots & 2v_m
\end{matrix}
$$

is positive definite ($v_j > 0$; for all j), the solution is a global minimum.

The final equation for the variance estimate is:

$$
(W.2) \quad C_{n+1} = \left\{ \sum_{j=0}^{m} \frac{y_j^2}{v_j} \sum_{j=0}^{m} \frac{1}{v_j} - \left(\sum_{j=0}^{m} \frac{y_j}{v_j} \right)^2 \right\}^{-2} \sum_{j=0}^{m} \left[y_j \left(\hat{\theta} \sum_{j=0}^{m} \frac{1}{v_j} - \sum_{j=0}^{m} \frac{y_j}{v_j} \right) + \left(\sum_{j=0}^{m} \frac{y_j^2}{v_j} - \hat{\theta} \sum_{j=0}^{m} \frac{y_j}{v_j} \right) \right]^2 \frac{1}{v_j}
$$

Now the equations ($W.1.a$) and ($W.2$) contain all that is necessary to implement the WF procedure.

C.   THE WF PROCEDURE

1. Find $\theta_{min} = min(\hat{\theta}_n, v_{n+1,1} \dots v_{n+1,m})$, $\theta_{max} = max(\hat{\theta}_n, v_{n+1,1} \dots v_{n+1,m})$.

2. Fix the actual search resolution R (depending on $T = [\theta_{min}, \theta_{max}]$).

3. Set up the grid with resolution R over the interval $(\theta_{min} - w, \theta_{max} + w)$.
    (Let $k = w \cdot R$)

4. Evaluate the likelihood function at each grid point.

5. Find the $\hat{\theta}_{n+1}$ by picking the "best" grid point solution of equation ($W.1.a$) in interval T using sum of $2k+1$ adjacent likelihood function values for integral approximation.
    (See Appendix B for details on implementation).

6. Compute C_{n+1} by equation ($W.2$).

D.   PRACTICAL CONSIDERATIONS

A number of issues related to implementing the WF procedure on a digital computer are listed below. These considerations were used in in simulation described in the next Chapter.

1. Grid setup

The exhaustive search over the interval $T = [\theta_{min}, \theta_{max}]$ may be performed over a pre-specified number of points to be checked or using a pre-specified search resolution.

The first approach is wasteful when T is small and the second approach is wasteful when T is large. Thus, it would seem reasonable to specify both a desired resolution and an upper bound on the maximum number of points to be checked, which remain constant over time.

2. Very dense observations

In the case when T is very short there may be two complications:

1) Because of resolution problems $\hat{\theta}_{n+1}$ may take exact value of θ_{min} or θ_{max}. In this case equation ($IV.2$) may force C_{n+1} equal to the corresponding v_j, which is not right.

2) Computing C_{n+1} by ($IV.2$) with y_j values very close together may yield underflow and division by zero.

In order to prevent these problems the following modification is suggested and was used in the simulations: if T is very short (say less than 3 times resolution) set $\hat{\theta}_{n+1}$ to be equal to midpoint of T and compute C_{n+1} using ($K.2$).

3. "Very small" numbers

When m is large, the likelihood function takes very low values. To avoid underflow, simply multiply it by a large constant. The constant that was used in the simulation is 1024^m.

One more modification was implemented to avoid arithmetic with very small numbers. If the exponent of the likelihood function first term was less than -100, the likelihood function was set to be equal zero. The meaning of the modification is that the procedure considers any target step larger than $\sim 7 \times$ (standard deviation) to be a practically an impossible step. (One must be careful with such modifications if large target jumps are possible, as could occur if the model involved Student-t θ-jumps).

26

IV. SIMULATION RESULTS

A. SIMULATION DESCRIPTION

A simulation experiment was performed on an IBM 3033 and 4381 at the Naval Postgraduate School. The programming language used was FORTRAN-77, and single precision was used. Normal and Chi-square random numbers were generated with the aid of IMSL library routines. Student-t random numbers were generated from normal and Chi-square random variables with appropriate degrees of freedom (see [Ref. 5]). Tracking error statistics were computed using the HISTGP subroutine from the NONIMSL library at NPS. Figures were generated using an IBM experimental graphics package. GRAFSTAT.

For each simulation replication the tracking sequence was performed for 100 steps , (i.e. $n+1 = 1,2,...,100$). The number of replications was 1000 in all cases.

The KF and WF procedures (with different values for w) were carried out using the same random numbers. (With normally distributed target steps and Student-t distributed measurement errors).

Statistics were collected on tracking errors $\hat{\theta}_{n+1} - \theta_{n+1}$ for $n+1 = 1,25,50,75,100$.

The simulation was performed for 6 cases:

- Case 1 -one observer. $\sigma_1 = 1$
- Case 2 -two observers. $\sigma_1 = \sigma_2 = 1$
- Case 3 -three observers. $\sigma_1 = \sigma_2 = \sigma_3 = 1$
- Case 4 -three observers. $\sigma_1 = \sigma_2 = 1;$ $\sigma_3 = 3$
- Case 5 -five observers. $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = 1$
- Case 6 -five observers. $\sigma_1 = \sigma_2 = \sigma_3 = 1;$ $\sigma_4 = \sigma_5 = 3$

In all cases the measurement errors of the observers have the Student-t distribution with $d = 3$ degrees of freedom and scale σ_j, and the target has normally distributed step sizes, with standard deviation $\tau = 1$. In all cases the WF procedure was performed for values $w = 0,1,2,3$. Excluding case 1, the desired grid resolution, DR, was chosen to be 0.1. In case 1, the desired resolution was equal to 0.05. Case 1 is most susceptible to problems resulting from a very short interval T described in the previous chapter. In all cases the number of grid points on the interval T was limited by upper bound of 50. (See Appendix B for details on resolution and grid setup.)

After fixing a grid, the likelihood function was evaluated at each grid point over the interval of length $T + 2w$. The integral was approximated by the sum of $2k + 1$ values of the likelihood function with k being the largest integer not exceeding w/R.

B. WF PERFORMANCE

Figures C.1 through C.6 display comparisons of the standard deviations[4] of the tracking error, $\hat{\theta}_{n+1} - \theta_{n+1}$, for the KF and the WF for cases 1 through 6, respectively. In all the cases, the WF performance is considerably better than the KF performance. See Appendix A for tabulated results.

Figures C.7 and C.8 display WF tracking error standard deviations with different values of w. The optimal w is somewhere in the vicinity of 1 or 2, but sensitivity to the exact value of w is low. See Appendix A for exact numbers.

C. SENSITIVITY ANALYSIS

For sensitivity analysis purposes, the tracking sequence was performed on each case with the same steps and measurements, but the WF procedure used the following (wrong) assumptions:

- $\sigma_j = 0.5$ or $\sigma_j = 1.5$ instead of $\sigma_j = 1.0$ or $\sigma_j = 3.0$ (underestimating the measurement variance by a factor of 4)

- $d_j = 6$ instead of $d_j = 3$ (underestimating the measurement variance by a factor of 2)

- $\tau = 0.5$ instead of $\tau = 1$ (underestimating the target step size variance by a factor of 4)

Additional runs were performed when target step size was a Student-t distributed variable having 3 degrees of freedom and variance equal to 1.

In all sensitivity runs the WF used $w = 2$ and the desired resolution was equal to 0.1. Also, the maximal number of points in the interval T was 50.

Figures C.9 through C.14 display WF performance under sensitivity analysis conditions for cases 1 through 6, respectively. Clearly the WF procedure is very robust

---

[4]All the results presented in terms of sample standard deviations, i.e. a square roots of the *unbiased* estimates of population variances. Means are not presented since, in all cases for both filters, mean error values were very small. RMS (square Roots of Mean Square error) values were also computed in the simulation. In all runs inspected (the vast majority from all runs) RMS values were were slightly smaller than the standard deviations values (due to division by N in RMS computation and division by $N - 1$ in standard deviations computation).

28

with respect to measurement error parameter estimation and less robust with respect to major alternations in target movement model. See Appendix A for tabulated results.

The strange behavior of the standard deviation plot with t-distributed target step sizes is explained by the following phenomena: in very rare cases when the estimate $\hat{\theta}_n$ and actual target position θ_{n+1} are extremely distant (the target makes a huge step), WF starts to "ignore" the incoming measurements, regarding them as outliers and sticks to the old estimate. In such situation, C_n is increasing steadily but slowly (by 1τ each time step - due to the wrong target model). After some time (which may be as many as 30 time steps), C_n becomes very large and WF stops disregarding observations and the estimate shifts toward the actual target position. (The last modification described in the previous Chapter was harmful in this situation.)

Such undesired behavior may be treated by ad hoc adjustments to the WF procedure. For example, after detecting a steady increase in C_n we may arbitrary set C_n equal to a large number and recalculate the last estimate. This kind of adjustment may be proper when actually implementing WF if there is concern about the normality of target step size. None of these adjustments were implemented in the simulation experiment. It is anticipated that if an outlier-productive model of the target motion were used, the effect noted would be automatically reduced, but to date no investigation has been made.

Another comparison was made between the WF and KF procedures by performing a simulation experiment with normally distributed measurement errors. The simulation was performed for cases 1a-6a. Cases 1a-6a correspond to cases 1-6, respectively, with normally distributed measurement errors having the same variance as t-distributed errors in the original cases. Once again the WF used $w = 2$ and the desired resolution was set to 0.1 with a maximum number of points in the interval T of 50.

Figures C.15 and C.16 display the comparison of WF and KF procedures for cases 1a-6a. It is evident that the advantage of KF over WF in cases 1a-6a is less than the superiority of WF over KF in cases 1-6. See Appendix A for tabulated results.

In order to see how much measurement error assumptions may be violated, several more runs were made using the Cauchy distribution (Student-t with 1 d.f.) for generating measurement errors (WF assumed 3 d.f. and the correct second

29

parameter σ). In general the tracking was steady. However in very extreme cases, when the interval T was extremely long and the old estimate $\hat{\theta}_n$ was not one of it's endpoints (actual resolution in such cases became extremely large), WF produced an unreasonable estimate and did not recover from the huge error.

To overcome the problem, two alterations were made to the simulation implementation:

- The Maximum number of points to check was set to 500.
- In the case when *all* grid points on the interval T corresponded to the integral having a value equal to 0, the new estimate was set to be equal to $\hat{\theta}_n$, the old estimate. The original implementation produced θ_{min} for an estimate in such cases. (See Appendix B for a better approach by altering grid setup.)

After making these two alterations, the runs were repeated. The tracking was steady with no special problems. The results are tabulated in Appendix A and may be considered very good.

A selected number of cases was chosen to perform a limited variability demonstration. Simulation runs were made with 5 different pairs of seeds (one for target steps and one for measurement errors) for 1,2,3 and 5 identical observers and WF using $w=2$. The results are displayed in Figure C.17. Generally, variability between different pairs of seeds does not exceed the variability between separated time steps for a single seed.

V. FINAL DISCUSSION

The WF procedure has a very general nature. It may be applied directly to any type of measurement error distribution. The distribution of measurement error must have finite variance and known density (even in tabular form). The WF filter performance with exotic distributions, using right or wrong distributional assumptions, needs to be investigated.

(To apply the WF procedure to models with non-normal target step size there must be an efficient way to evaluate the conditional likelihood function, or else additional assumptions have to be made regarding distribution of the estimate $\hat{\theta}_n$ before and after movement of the target.)

When applied to models with normally distributed measurement errors (and normally distributed target step sizes), the WF procedure (with any w) is theoretically identical to the KF procedure. In practice, however, there might be some difference in performance due to finite resolution.

WF has a built-in tuning parameter: the value of w. The optimal value for w is definitely not zero, but the sensitivity of the procedure to the exact value of w is low. The best values of w seem to be in the range $(\tau, 2\tau)$, where τ is the standard deviation of the target step size.

Instead of using w as a static tuning parameter where it is kept constant, w may be computed dynamically, based on the latest observations. For example, $w = T/4$ constrained to $w > 0.3$ and $w < 4$ and the usual procedure for grid setup (see Appendix B) produced very good results for all cases, but was not superior[5] to the results obtained with constant w. (Dynamic w results were not represented in the previous Chapter).

Another modification to be explored, if it can be justified by operational reasons, is use of a non-symmetric interval in equation $(IV.1.a)$, i.e. performing integration with lower limit $\theta_{n+1} - w_1$ and upper limit $\theta_{n+1} + w_2$.

Equation $(IV.2)$ may be used not only as part of WF procedure, but as part of any "good" procedure specifying $\hat{\theta}_{n+1}$ and looking for C_{n+1}.

---

[5]Optimal rule for dynamic w calculation may be discovered by further research.

The most noticable difference between the KF and the WF with respect to C_n estimation is that in the KF case C_n depends only on n, converging to some value as n increases, and in the WF case C_n is more dynamic, depending on the actual measurement values. The dynamic behavior of C_n reflects the variable amounts of information received at each measurement step. This phenomenon gives an "inside" performance measure on how well the filter is doing. None of the cases checked by simulation required corrective actions due to abnormal C_n behavior (i.e. steady increase or large jumps), except when target step sizes were Student-t distributed.

A limited inspection of C_n values calculated by the WF procedure indicates that the C_n values computed are somewhat larger than the actual variances of tracking errors. It may conjectured that there is room for improvement in the WF procedure. On the other hand, forcing C_n close to its true value (known from previous simulations) produces good but not superior results to those obtained in the regular way, using equation (W.2). This suggests that the WF procedure is fairly robust with respect to modest changes in the C_n computation process. (A constant value for C_n may be used if there is sufficient knowledge of the environment, i.e. the C_n values that should be obtained are approximately known and no large jumps in C_n values are expected).

Extending the WF procedure to the multidimensional case does not seem to pose conceptual problems. Equation (W.1.a) may be replaced by an integral over a k-dimensional rectangle. The analog to equation (W.2) may be obtained by minimizing the determinant of the covariance matrix. However, the computational burden of solving equation (W.1.a) in the k-dimensional case and minimizing the determinant may be excessive. To actually implement the WF procedure in the multidimensional environment would require an efficient way to solve equation (W.1.a) and to minimize the determinant of the covariance matrix.

APPENDIX A
TABULATED RESULTS

This Appendix contains tables of tracking error standard deviations. The tables are organized by observers configuration.

1. Table 1: One observer.
2. Table 2: Two identical observers.
3. Table 3: Three identical observers.
4. Table 4: Three observers, one of whom is 3 times less accurate than the others.
5. Table 5: Five identical observers.
6. Table 6: Five observers, two of whom is 3 times less accurate than the others.

The following notation is used for specifying target step size and measurement error distributions:

- $N(v)$ - normal with mean 0 and variance v
- $t(d,s)$ - Student-t with d degrees of freedom and scale parameter $\sigma = s$

When two types of observers are involved, the parameters are given for the more accurate ones.

All cases above the double line use precisely the same target steps and measurement errors; only the filters are different.

Filter notations:

- KF   Kalman filter using correct variance
- WF $w=i$ W-filter using 3 d.f., $w=i$ and correct variance
- WF$w=i,\frac{1}{2}\sigma$ W-filter using 3 d.f., $w=i$ and assuming the scale of the t is $\frac{1}{2}\sigma$ rather than the true parameter σ for each observer
- Wf$w=i,2d$ W-filter using $w=i$ and assuming that the Student-t degrees of freedom are $2d$ rather than the true value of d
- WF$w=1,\frac{1}{2}\tau$ W-filter the correct measurement parameters and assuming the standard deviation of the target step size is $\frac{1}{2}\tau$ rather than the true value of τ

TABLE 1
ONE OBSERVER

Mvmnt step	Measur. Error	FILTER	Track Length				
			1	25	50	75	100
N(1)	t(3,1)	KF	0.830	1.108	1.079	1.121	1.117
"	"	WF $w=0$	0.842	1.160	1.082	1.176	1.094
"	"	WF $w=1$	0.832	1.096	1.037	1.124	1.048
"	"	WF $w=2$	0.835	1.024	0.998	1.040	0.993
"	"	WF $w=3$	0.837	1.033	1.021	1.060	1.031
"	"	WF $w=2,\tfrac{1}{2}\sigma$	0.846	1.017	1.003	1.042	0.993
"	"	WF $w=2,2d$	0.821	1.013	0.993	1.038	0.998
"	"	WF $w=2,\tfrac{1}{2}\tau$	0.915	1.146	1.192	1.209	1.221
t(3,.58) N(1)	t(1,1)	WF $w=2$	0.853	1.632	1.702	1.141	1.125
		WF $w=2$	0.884	1.349	1.445	1.362	1.406
N(1)	N(3)	KF	0.857	1.203	1.166	1.124	1.131
		WF $w=2$	0.871	1.250	1.209	1.185	1.192

TABLE 2
TWO OBSERVERS

Mvmnt step	Measur. Error	FILTER	Track Length				
			1	25	50	75	100
N(1)	t(3,1)	KF	0.735	0.891	0.859	0.888	0.879
"	"	WF $w=0$	0.672	0.801	0.757	0.784	0.771
"	"	WF $w=1$	0.670	0.792	0.746	0.777	0.766
"	"	WF $w=2$	0.682	0.787	0.743	0.777	0.769
"	"	WF $w=3$	0.713	0.820	0.779	0.816	0.808
"	"	WF $w=2,\tfrac{1}{2}\sigma$	0.684	0.786	0.746	0.776	0.767
"	"	WF $w=2,2d$	0.683	0.796	0.749	0.784	0.775
"	"	WF $w=2,\tfrac{1}{2}\tau$	0.835	0.909	0.902	0.933	0.917
t(3,.58) N(1)	t(1,1)	WF $w=2$	0.749	1.295	1.499	0.838	0.857
		WF $w=2$	0.813	1.004	1.080	1.006	1.023
N(1)	N(3)	KF	0.750	0.924	0.955	0.908	0.914
		WF $w=2$	0.778	0.959	0.989	0.946	0.957

TABLE 3
THREE IDENTICAL OBSERVERS

Mvmnt step	Measur. Error	FILTER	Track Length				
			1	25	50	75	100
N(1)	t(3,1)	KF	0.663	0.774	0.748	0.746	0.778
-"-	-"-	WF $w=0$	0.587	0.682	0.660	0.633	0.667
-"-	-"-	WF $w=1$	0.586	0.673	0.653	0.632	0.658
-"-	-"-	WF $w=2$	0.600	0.669	0.654	0.645	0.662
-"-	-"-	WF $w=3$	0.640	0.694	0.687	0.680	0.707
-"-	-"-	WF$w=2,\tfrac{1}{2}\sigma$	0.602	0.668	0.658	0.646	0.662
-"-	-"-	WF$w=2,2d$	0.601	0.674	0.659	0.652	0.673
-"-	-"-	WF$w=2,\tfrac{1}{2}\tau$	0.778	0.775	0.802	0.793	0.820
t(3,.58) N(1)	-"- t(1,1)	WF $w=2$ WF $w=2$	0.661 0.750	1.073 0.843	1.411 0.890	0.720 0.859	0.709 0.836
N(1) -"-	N(3) -"-	KF WF $w=2$	0.723 0.745	0.815 0.849	0.789 0.809	0.794 0.826	0.768 0.796

TABLE 4
THREE OBSERVERS, ONE 3 TIMES WORSE

Mvmnt step	Measur. Error	FILTER	Track Length				
			1	25	50	75	100
N(1)	t(3,1)	KF	0.721	0.887	0.866	0.871	0.874
-"-	-"-	WF $w=0$	0.642	0.793	0.766	0.762	0.769
-"-	-"-	WF $w=1$	0.641	0.778	0.758	0.754	0.760
-"-	-"-	WF $w=2$	0.659	0.768	0.736	0.716	0.729
-"-	-"-	WF $w=3$	0.696	0.795	0.782	0.784	0.802
-"-	-"-	WF$w=2,\tfrac{1}{2}\sigma$	0.606	0.796	0.774	0.778	0.773
-"-	-"-	WF$w=2,2d$	0.659	0.775	0.756	0.762	0.768
-"-	-"-	WF$w=2,\tfrac{1}{2}\tau$	0.825	0.878	0.905	0.897	0.923
t(3,.58) N(1)	-"- t(1,1)	WF $w=2$ WF $w=2$	0.740 0.807	1.079 0.963	1.455 0.989	0.809 0.937	0.805 0.958
N(1) -"-	N(3) -"-	KF WF $w=2$	0.770 0.798	0.914 0.958	0.908 0.937	0.900 0.948	0.878 0.914

TABLE 5
FIVE IDENTICAL OBSERVERS

Mvmnt step	Measur. Error	FILTER	Track Length				
			1	25	50	75	100
N(1) -"- -"- -"-	t(3,1) -"- -"- -"-	KF WF $w=0$ WF $w=1$ WF $w=2$ WF $w=3$	0.565 0.476 0.480 0.496 0.547	0.656 0.536 0.535 0.537 0.577	0.616 0.528 0.524 0.530 0.567	0.632 0.534 0.532 0.545 0.588	0.603 0.512 0.507 0.513 0.588
-"- -"- -"-	-"- -"- -"-	WF$w=2,\tfrac{1}{2}\sigma$ WF$w=2,2d$ WF$w=2,\tfrac{1}{2}\tau$	0.508 0.497 0.681	0.555 0.542 0.629	0.546 0.534 0.640	0.560 0.552 0.677	0.535 0.517 0.631
t(3..58) N(1)	-"- t(1,1)	WF $w=2$ WF $w=2$	0.524 0.671	1.520 0.668	1.396 0.689	0.552 0.683	0.527 0.655
N(1) -"-	N(3) -"-	KF WF $w=2$	0.618 0.635	0.657 0.695	0.663 0.686	0.650 0.681	0.655 0.675

TABLE 6
FIVE OBSERVERS, TWO 3 TIMES WORSE

Mvmnt step	Measur. Error	FILTER	Track Length				
			1	25	50	75	100
N(1) -"- -"- -"-	t(3,1) -"- -"- -"-	KF WF $w=0$ WF $w=1$ WF $w=2$ WF $w=3$	0.645 0.577 0.567 0.586 0.617	0.787 0.647 0.641 0.645 0.671	0.739 0.641 0.633 0.639 0.670	0.751 0.647 0.645 0.654 0.697	0.700 0.599 0.601 0.611 0.649
-"- -"- -"-	-"- -"- -"-	WF$w=2,\tfrac{1}{2}\sigma$ WF$w=2,2d$ WF$w=2,\tfrac{1}{2}\tau$	0.606 0.584 0.760	0.669 0.644 0.745	0.657 0.644 0.761	0.677 0.664 0.793	0.636 0.616 0.749
t(3..58) N(1)	-"- t(1,1)	WF $w=2$ WF $w=2$	0.635 0.802	0.720 0.801	1.435 0.830	0.690 0.828	0.630 0.852
N(1) -"-	N(3) -"-	KF WF $w=2$	0.694 0.724	0.771 0.824	0.790 0.817	0.782 0.827	0.772 0.803

APPENDIX B
DETAILS ON IMPLEMENTATION

In this Appendix the implementation of the simulation experiment is discussed. (There are many different ways to implement the WF procedure.) In particular the implementation of the grid setup and search for maximal integral over interval with length $2w$ will be explained.

1. GRID SETUP

Two parameters (excluding w) control the grid setup:

1) DR - the desired resolution (usually $DR = 0.1$)

2) NMAX - the maximal number of points to search (usually $NMAX = 50$)

The first step is to find interval $T = [\theta_{min}, \theta_{max}]$, where

- $\theta_{min} = min(\hat{\theta}_n, y_{n+1,1}, \ldots, y_{n+1,m})$,

- $\theta_{max} = max(\hat{\theta}_n, y_{n+1,1}, \ldots, y_{n+1,m})$.

After finding T, the desired number of search points, n is computed: $n = (T/DR) + 1$. If $n > NMAX$, then n is set to be equal to NMAX.

In the next step, the actual resolution R is computed: $R = T/n$.

After fixing R, the number of points on interval w, k, is computed: $k = [w/R]$. (If w is computed dynamically its value may change).

Now the grid is fixed to have total number of $n + 2k + 1$ equally (R) spaced points having $n + 1$ points on interval T itself (one on each end).

The likelihood function values are evaluated for each grid point and stored in an array. (In the same order that the grid points appear on the real number line).

A more economical way to set up the grid would be as follows:

First, evaluate the likelihood function at the endpoints of interval T. If the likelihood is positive at both ends, proceed as described above. If the likelihood is essentially equal 0 on one or two endpoints, then eliminate the corresponding observation, recompute interval T, and repeat the procedure. (This was not implemented in the simulation experiment.)

2.   SEARCH FOR MAXIMAL INTEGRAL

Once all likelihood values are stored in the array, the search for the maximal integral was implemented in the following way.

Initiation step: compute sum of first $2k+1$ values from the array. This sum is the integral approximation corresponding to θ_{min}. This is the first candidate for $\hat{\theta}_{n+1}$.

Search loop: advance on the array each time adding a new element to the sum and subtracting the oldest one. If the new sum is larger than the largest among the old ones, take the corresponding grid point as the best candidate so far and record the largest sum.

After $n+1$ repetitions the procedure gives the best candidate among the $n+1$ grid points on interval T (and the corresponding integral value).

This implementation approach was chosen mainly because of its algorithmic simplicity. Adding and subtracting REAL numbers repetitively might introduce some round-off error. This is not a real problem in present case because:

- 50 additions and subtractions is not a very large number.
- The interest is in the grid point, not the corresponding integral value.
- In close cases when round-off error is reversing, the test can't be sure anyway because of finite resolution.

APPENDIX C
COLLECTION OF FIGURES

In this Appendix seventeen Figures presenting the WF performance (compared with the KF performance) are collected.

- Figures C.1 through C.6 present WF vs KF comparison for cases 1 through 6 respectively. WF performance is given for values of $w = 0,1,2,3$. The cases correspond to following observer configurations:
  - One observer
  - Two identical observers
  - Three identical observers
  - Three observers, one of whom 3 times less accurate than the others
  - Five identical observers
  - Five observers, two of whom 3 times less accurate than the others
- Figures C.7 and C.8 present WF performance with with different values of w
  - Figure C.7 - cases 1,2,3
  - Figure C.8 - cases 4,5,6
- Figures C.9 through C.14 present sensitivity of WF performance with respect to model violations. Figures C.9 through C.14 correspond to case 1 through case 6 respectively.
- Figures C.15 and C.16 present comparison of WF and KF performances for normally distributed measurement errors. Comparison presented for cases 1a-6a. Cases 1a-6a correspond to cases 1-6, but have normally distributed measurement errors with the same variance as Student-t distributed errors in original cases.
  - Figure C.15 presents cases 1a-3a
  - Figure C.16 presents cases 4a-6a
- Figure C.17 displays variability of the simulation using different seeds. Five pairs of seeds are used. Variability is presented for cases 1,2,3,5.
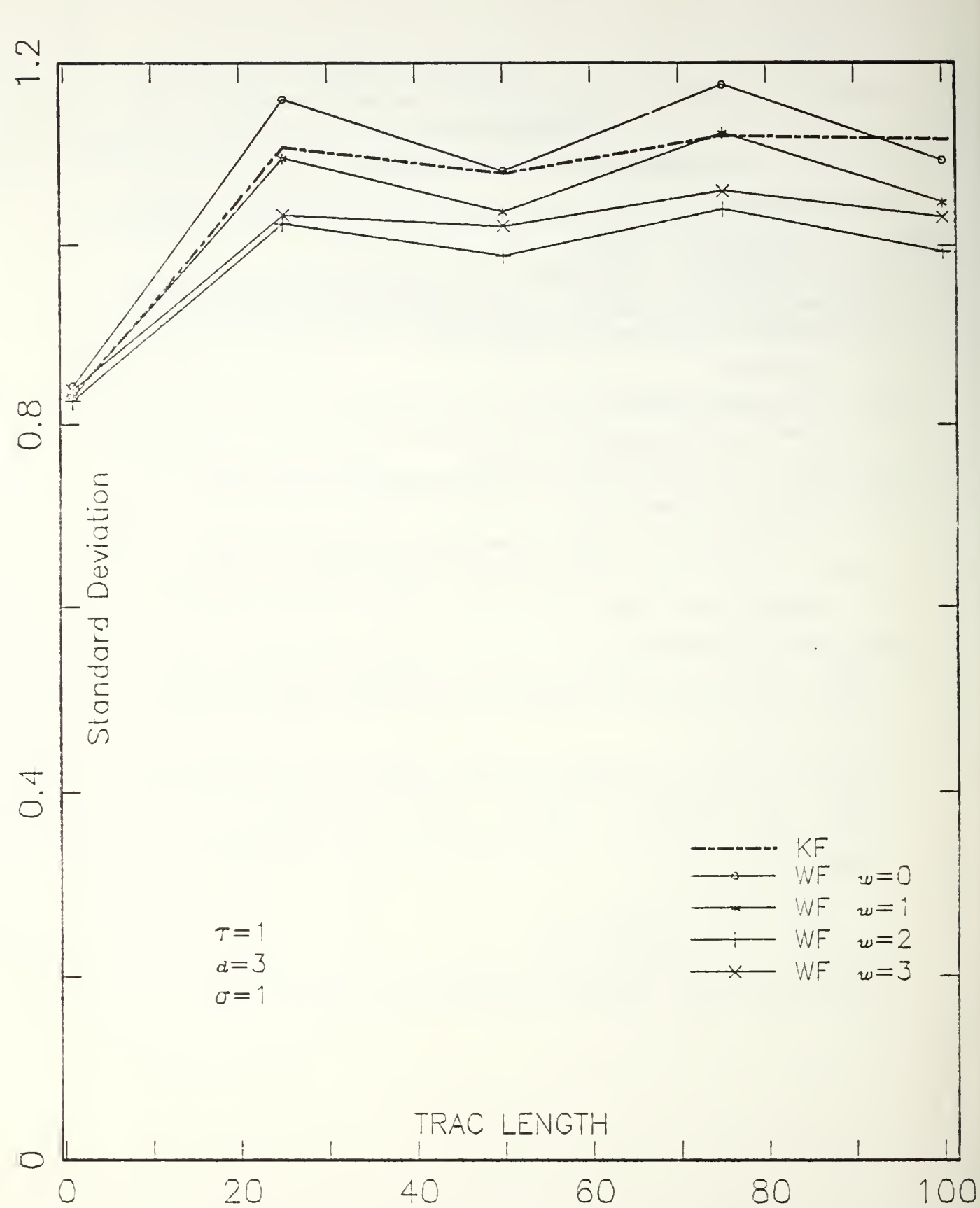
Figure C.1 WF vs KF comparison for Case 1 - One observer.

40
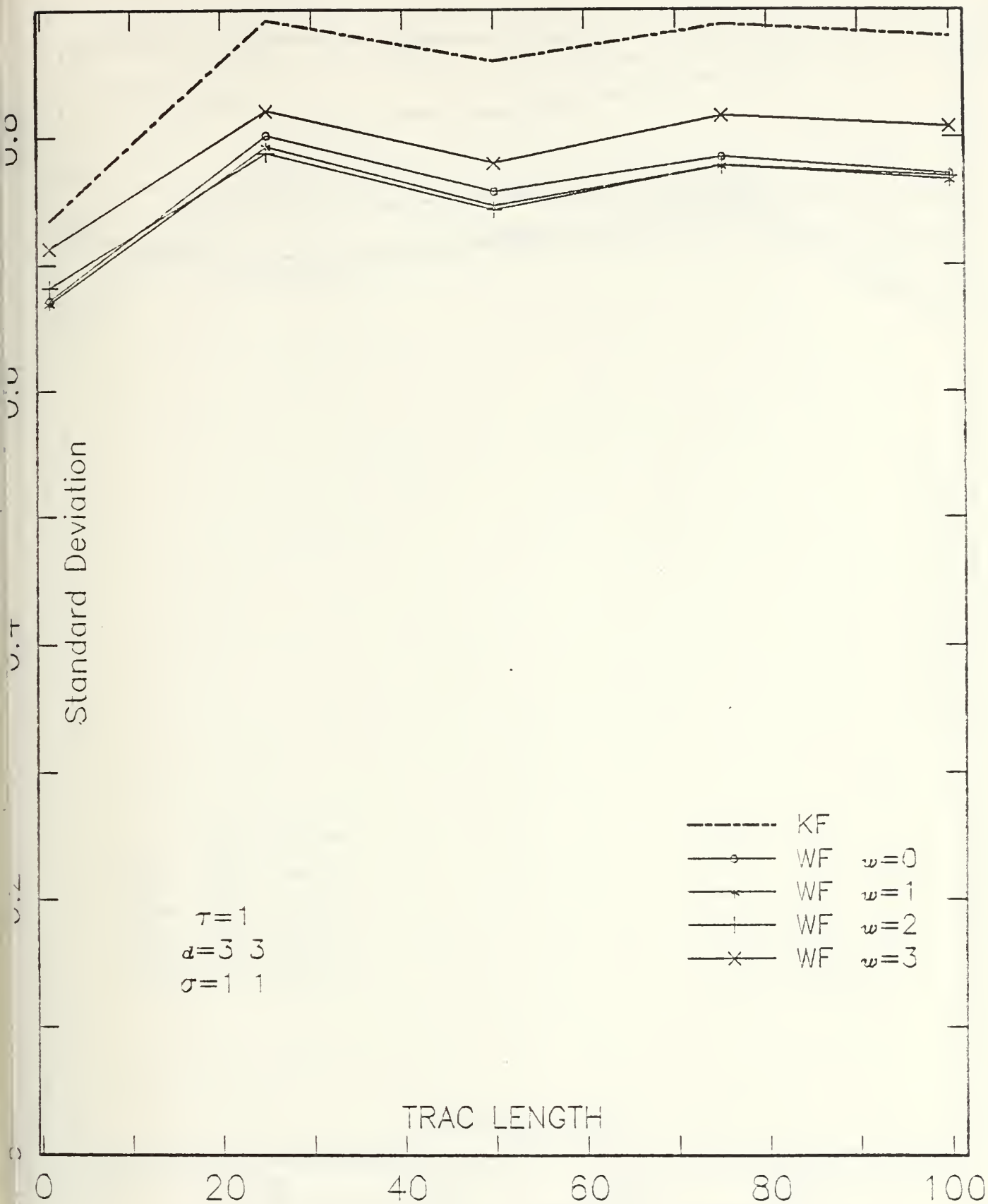
Figure C.2   WF vs KF comparison for Case 2 - Two Identical observers.
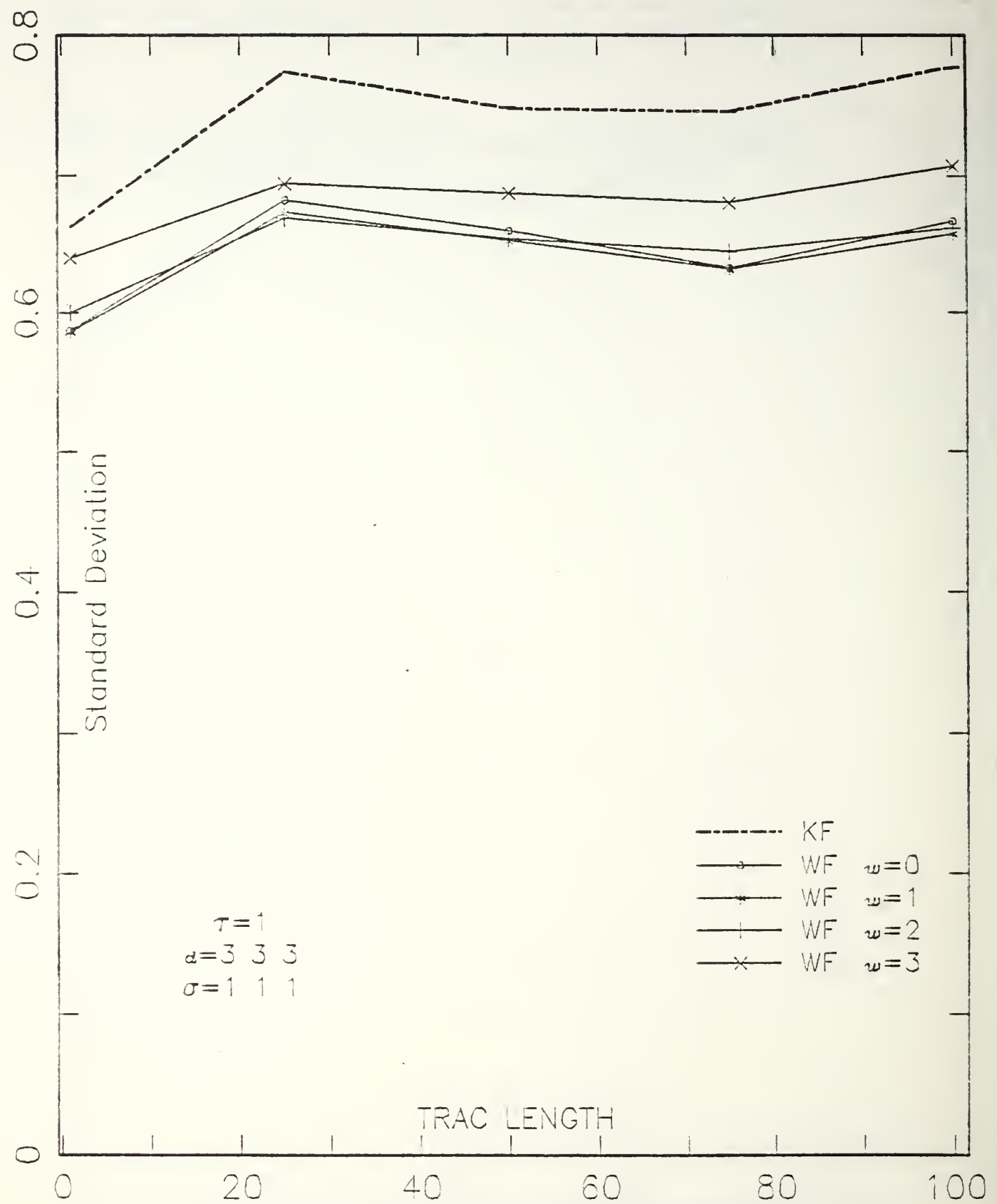
Figure C.3   WF vs KF comparison for Case 3 - Three Identical observers.

42
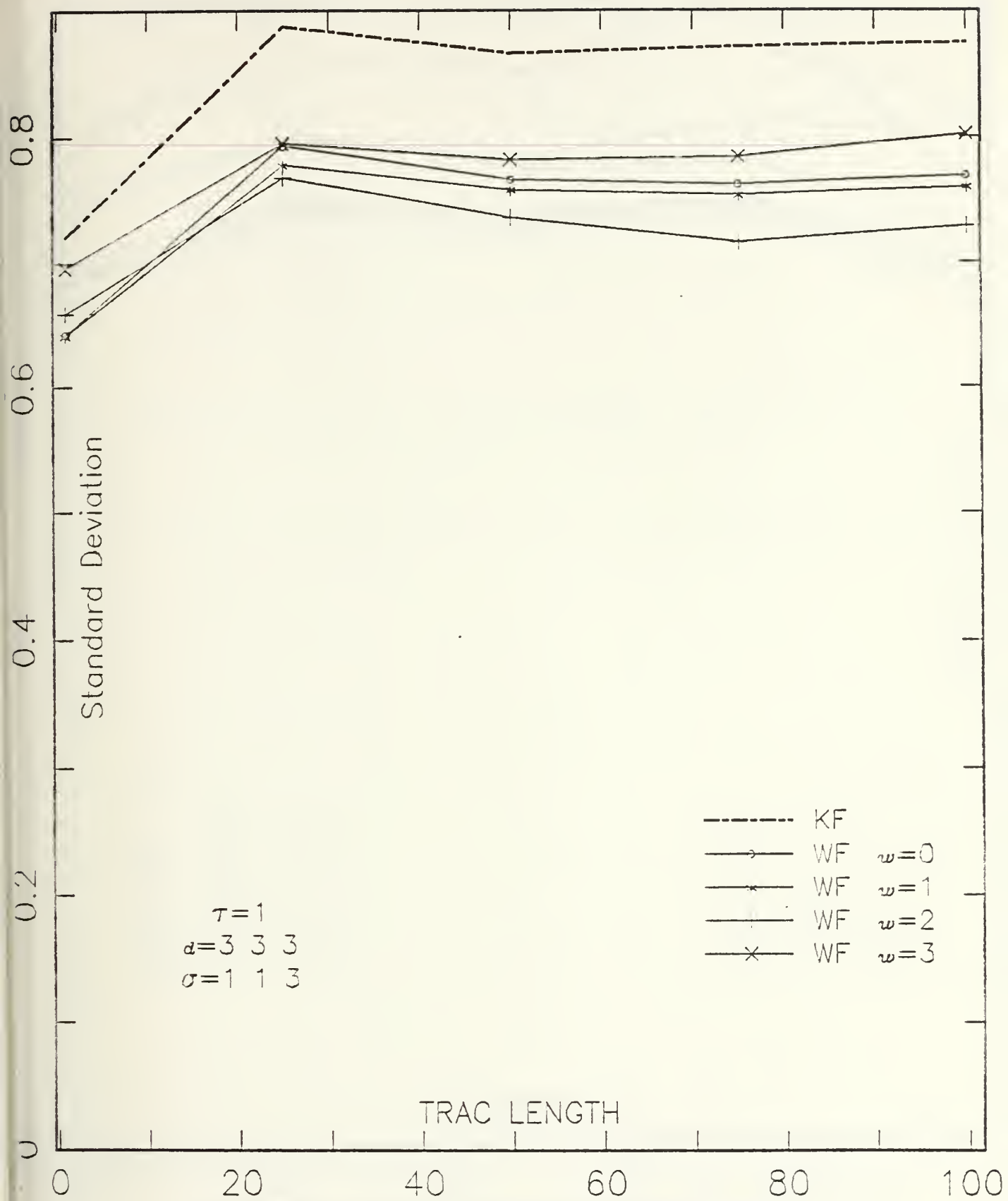
Figure C.4   WF vs KF comparison for Case 4 - Three Different observers.

43
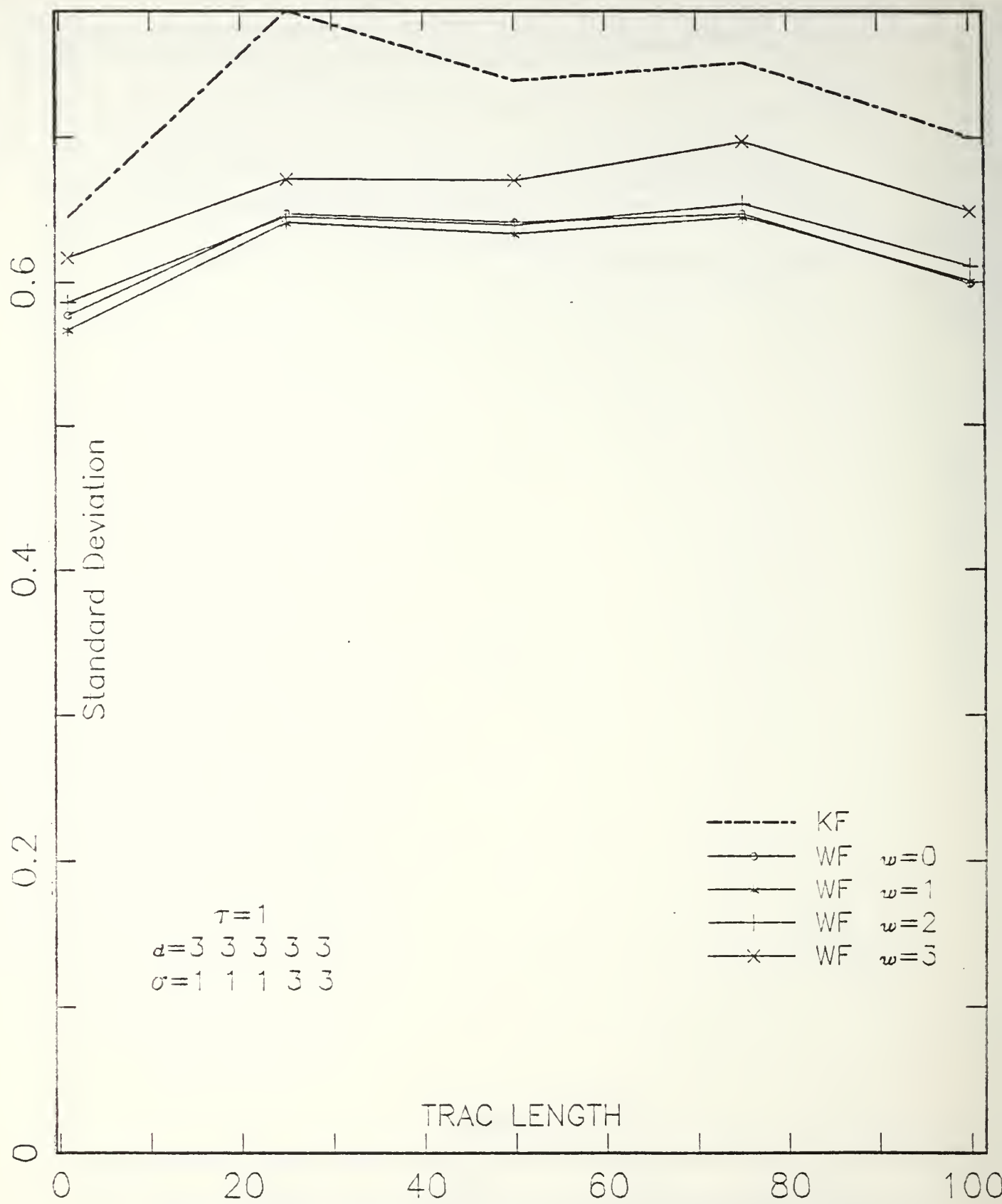
Figure C.5   WF vs KF comparison for Case 5 - Five Identical observers.

44
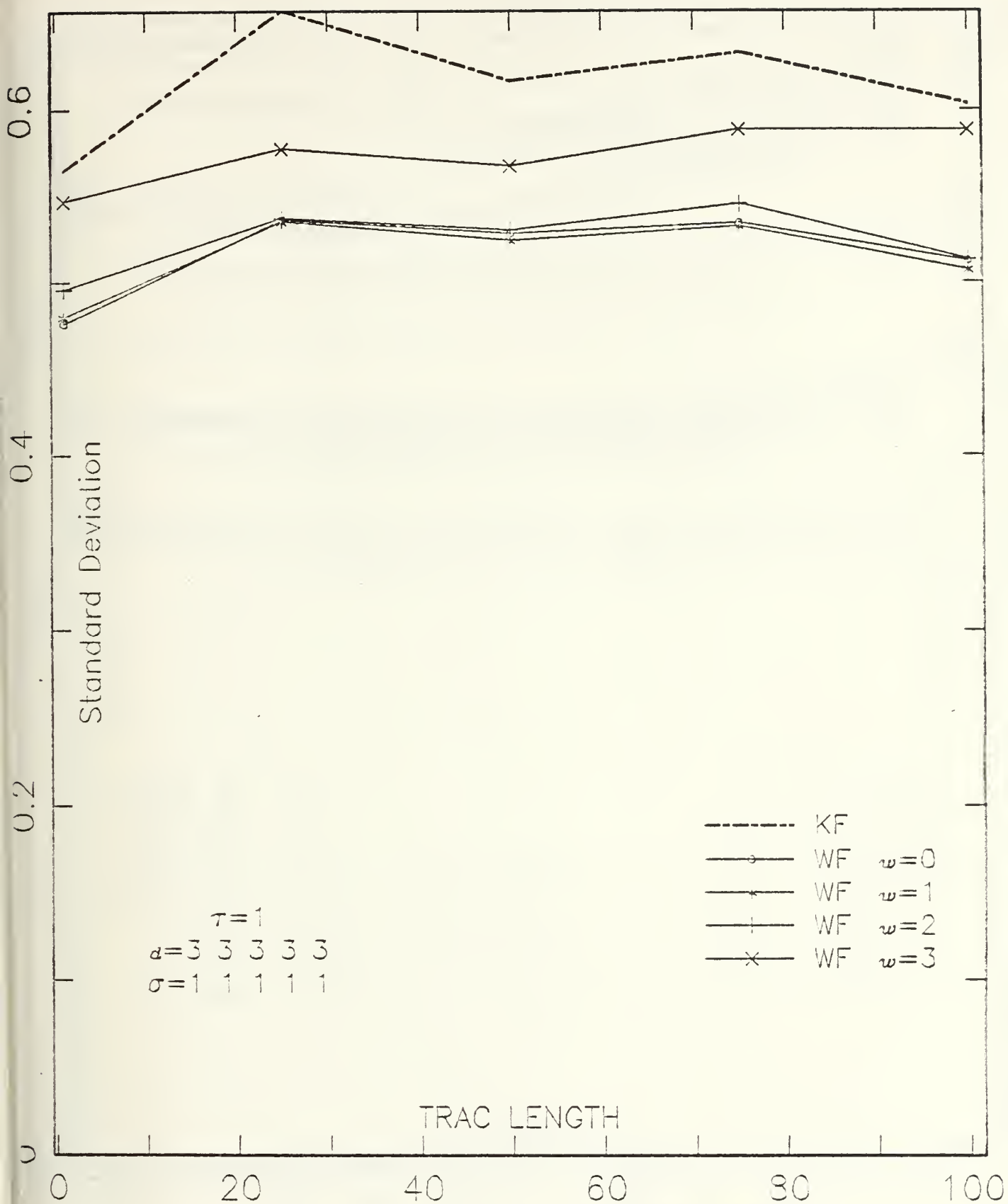
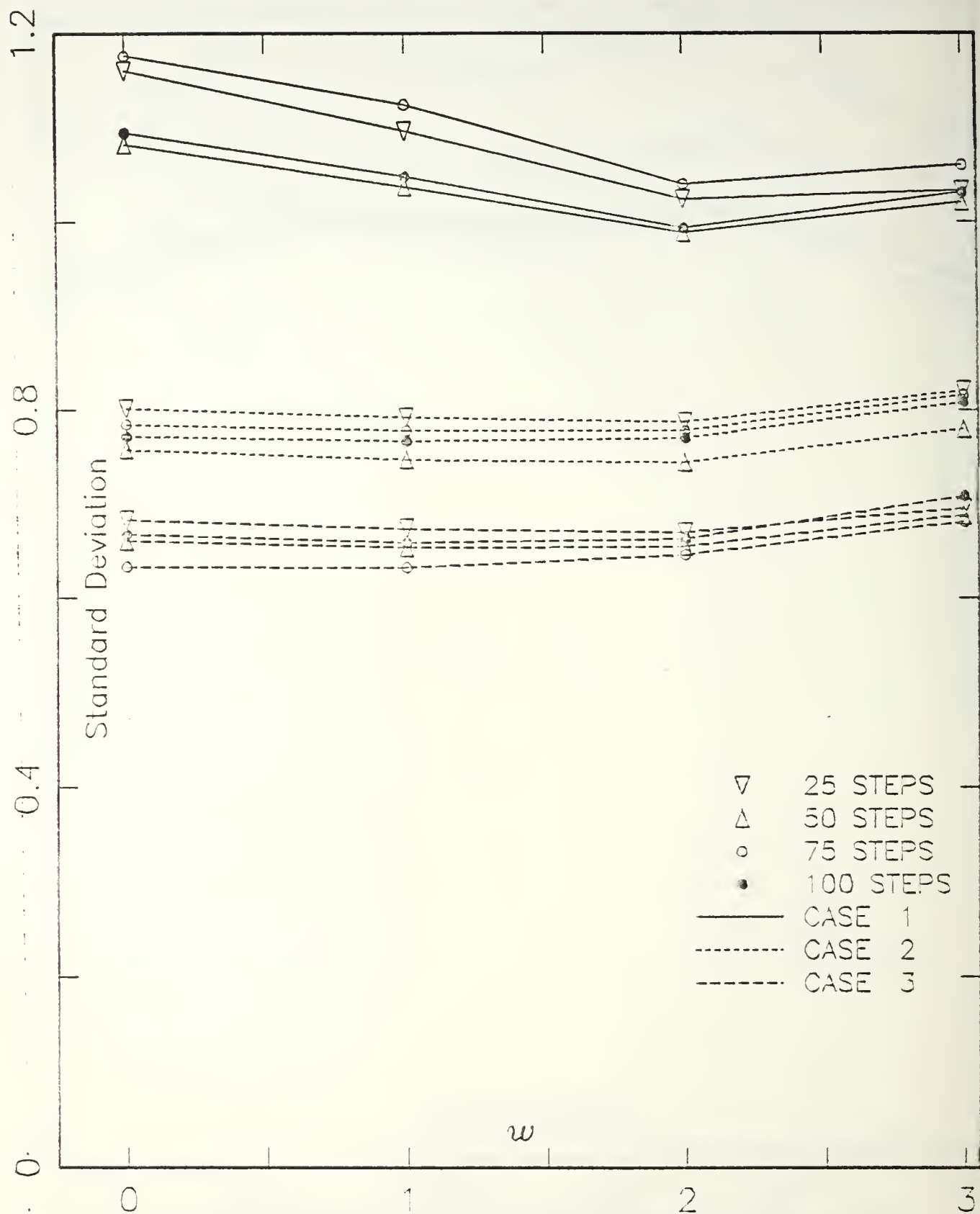Figure C.6   WF vs KF comparison for Case 6 - Five different observers.
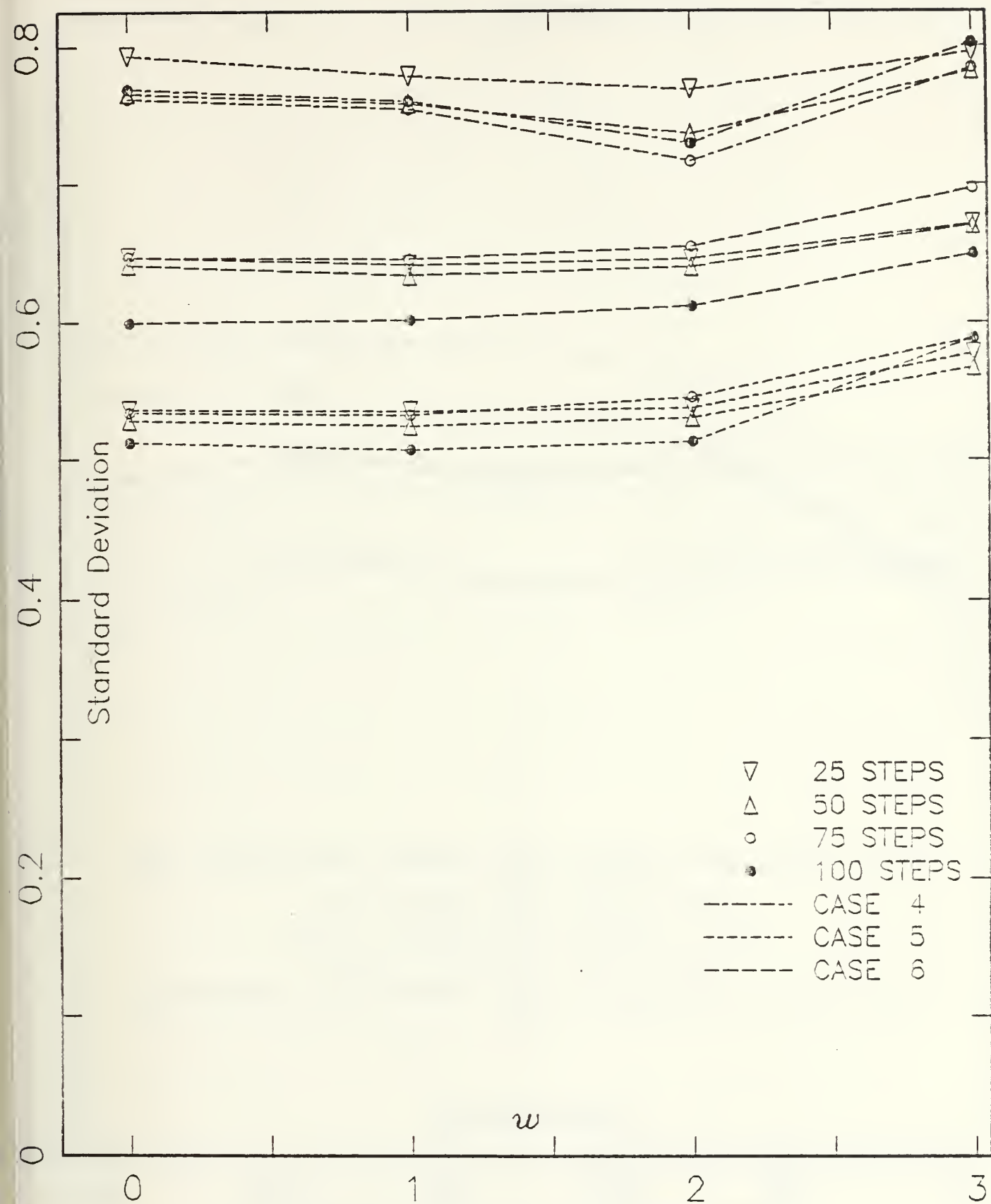
45

Figure C.7 Sensitivity to *w* for Cases 1-3.

46

Figure C.8   Sensitivity to *w* for Cases 4-6.
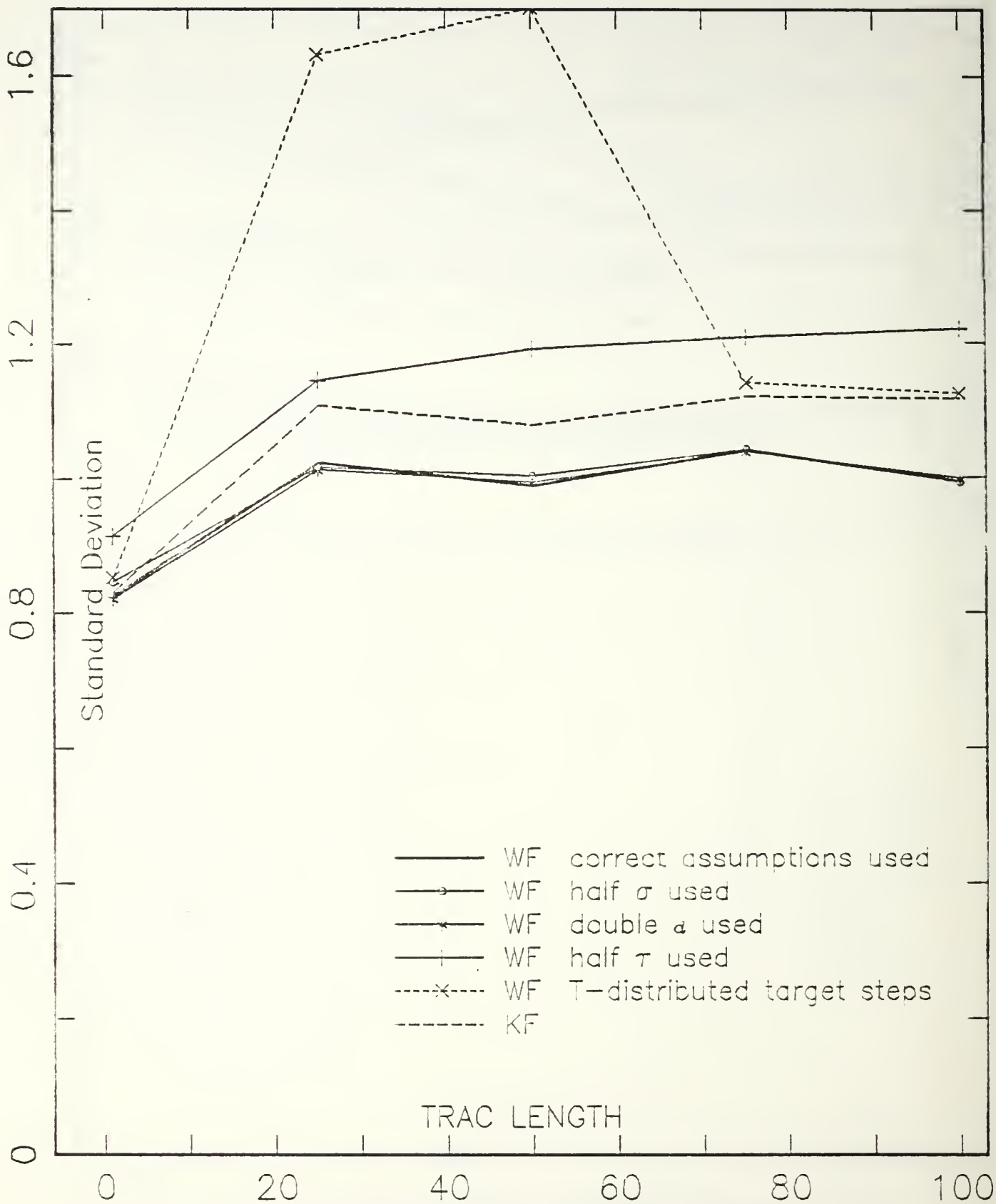
Figure C.9   WF Sensitivity Case 1 - One observer.

48
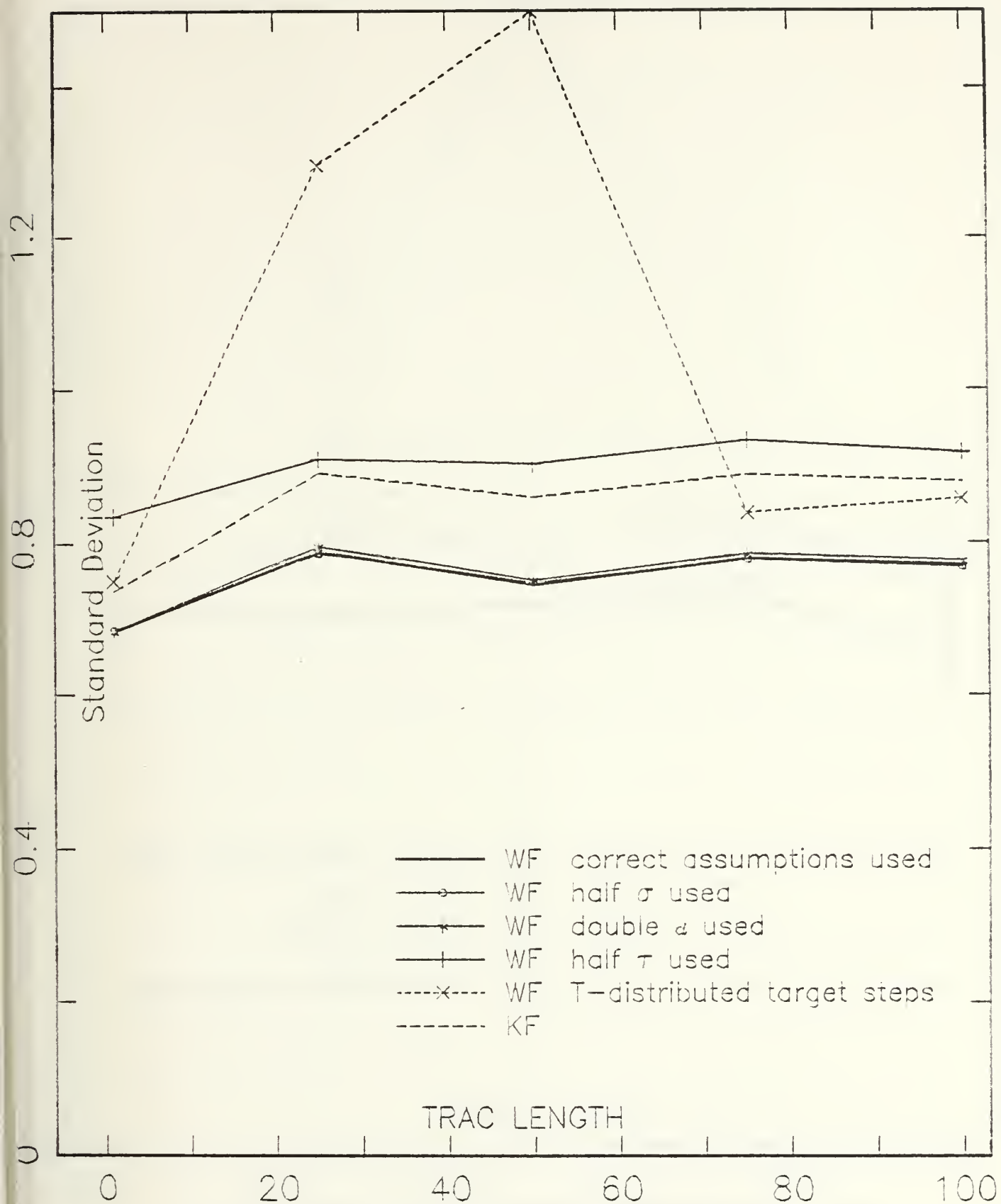
Figure C.10   WF Sensitivity Case 2 - Two Identical observers.

49
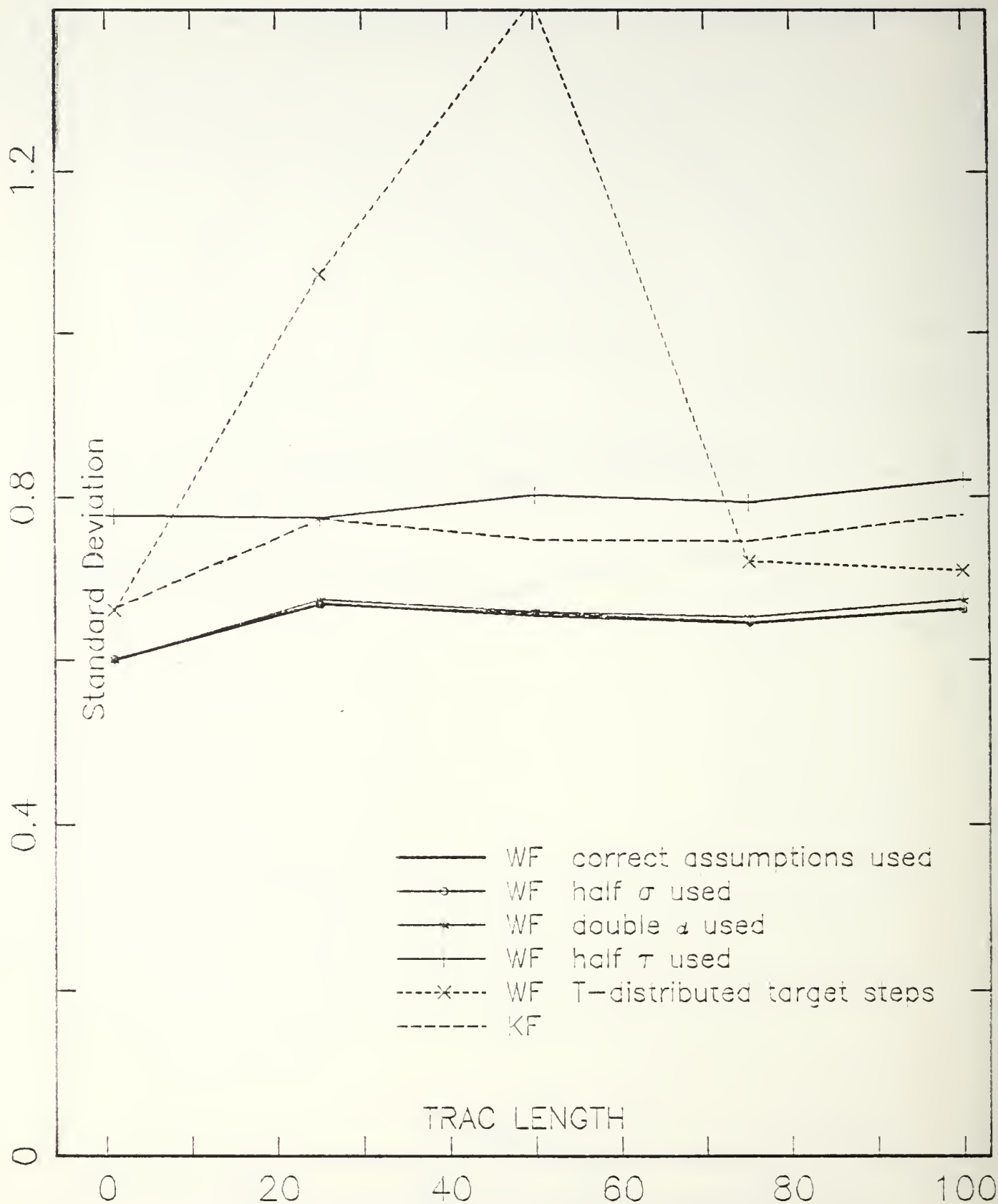
Figure C.11   WF Sensitivity Case 3 - Three Identical observers.

50
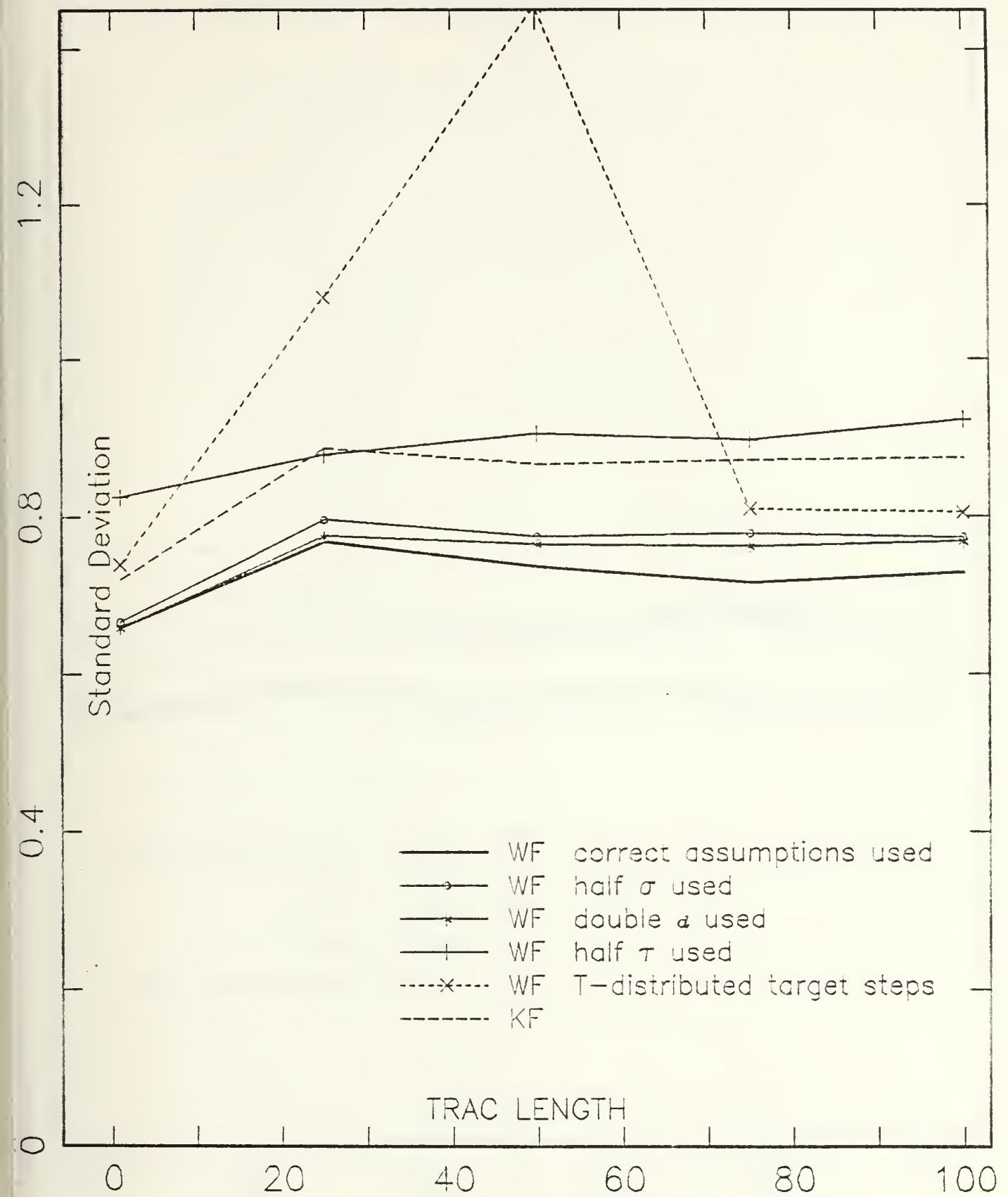
Figure C.12 WF Sensitivity Case 4 - Three Different observers.

51
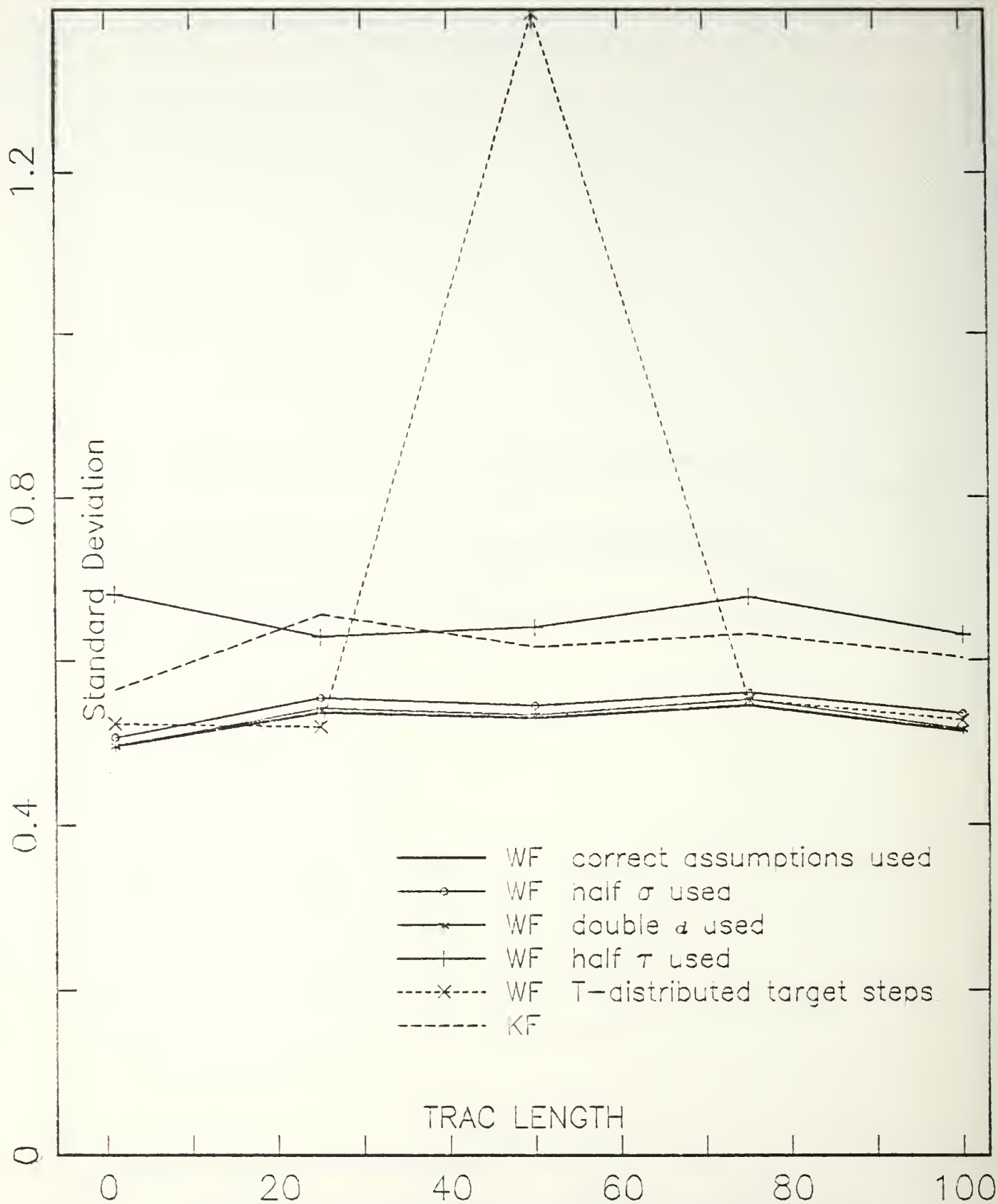
Figure C.13 WF Sensitivity Case 5 - Five Identical observers.
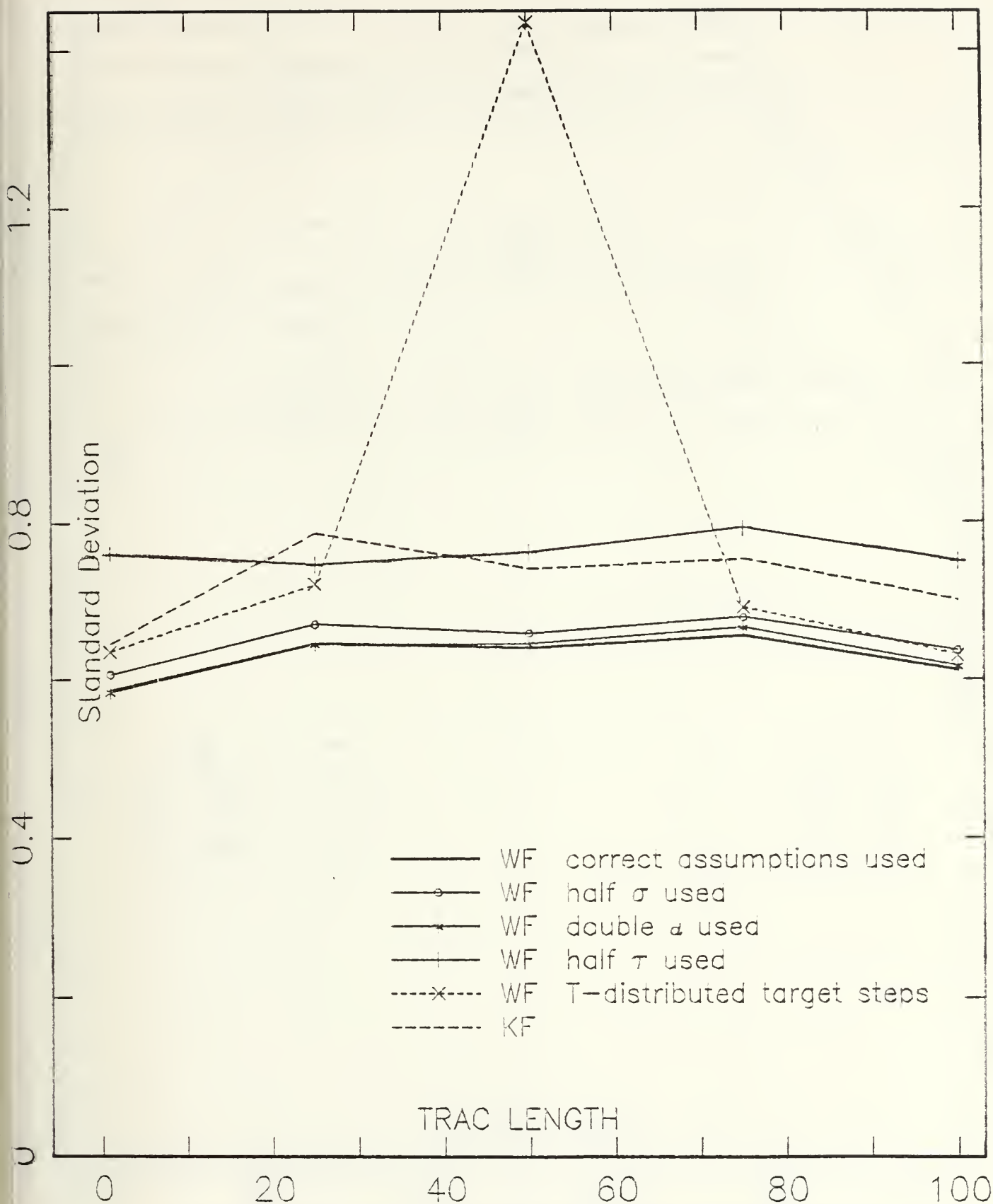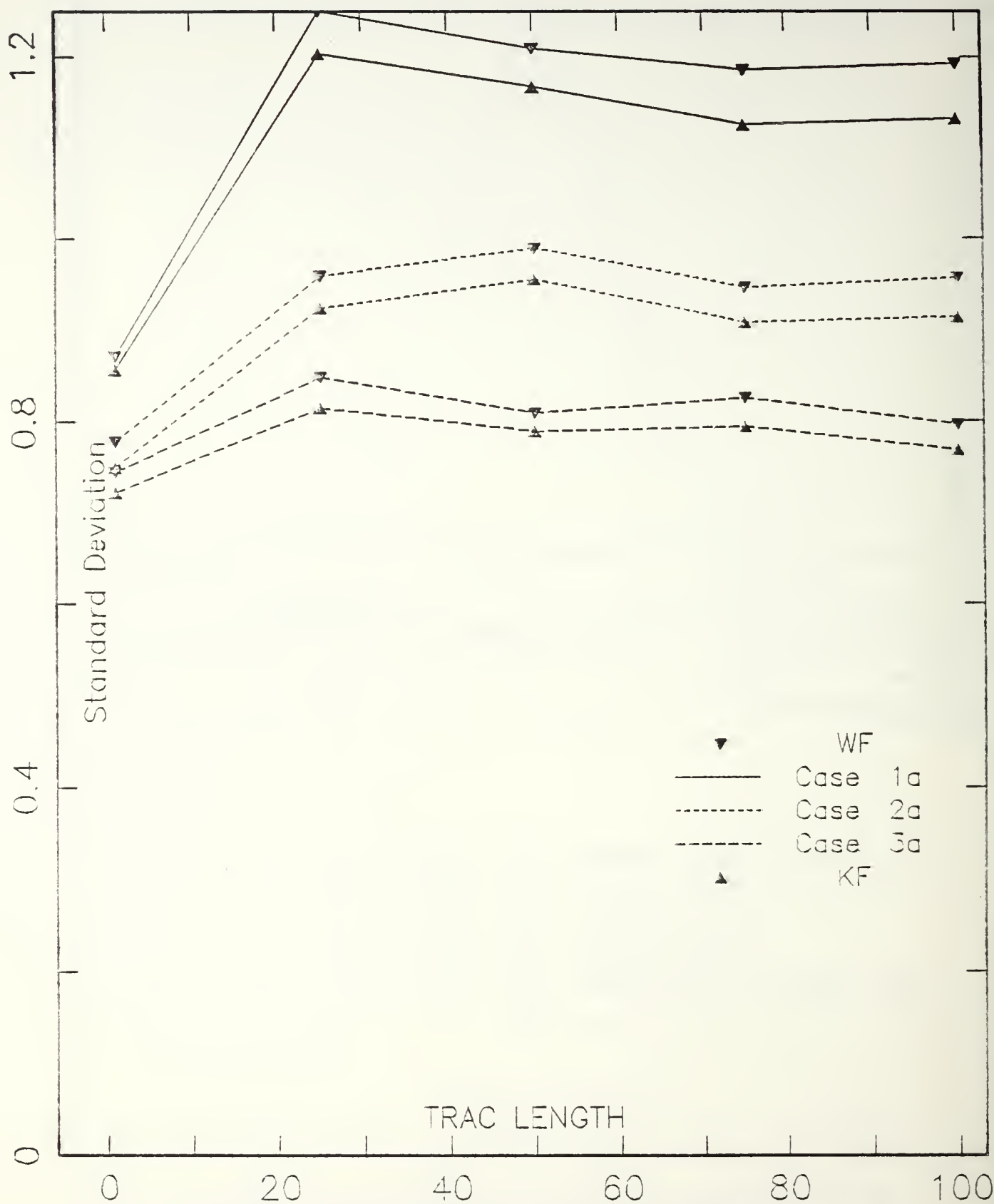
Figure C.14 WF Sensitivity Case 6 - Five Different observers.

53

Figure C.15   WF vs KF comparison for Case 1a-3a - Normal measurement errors.
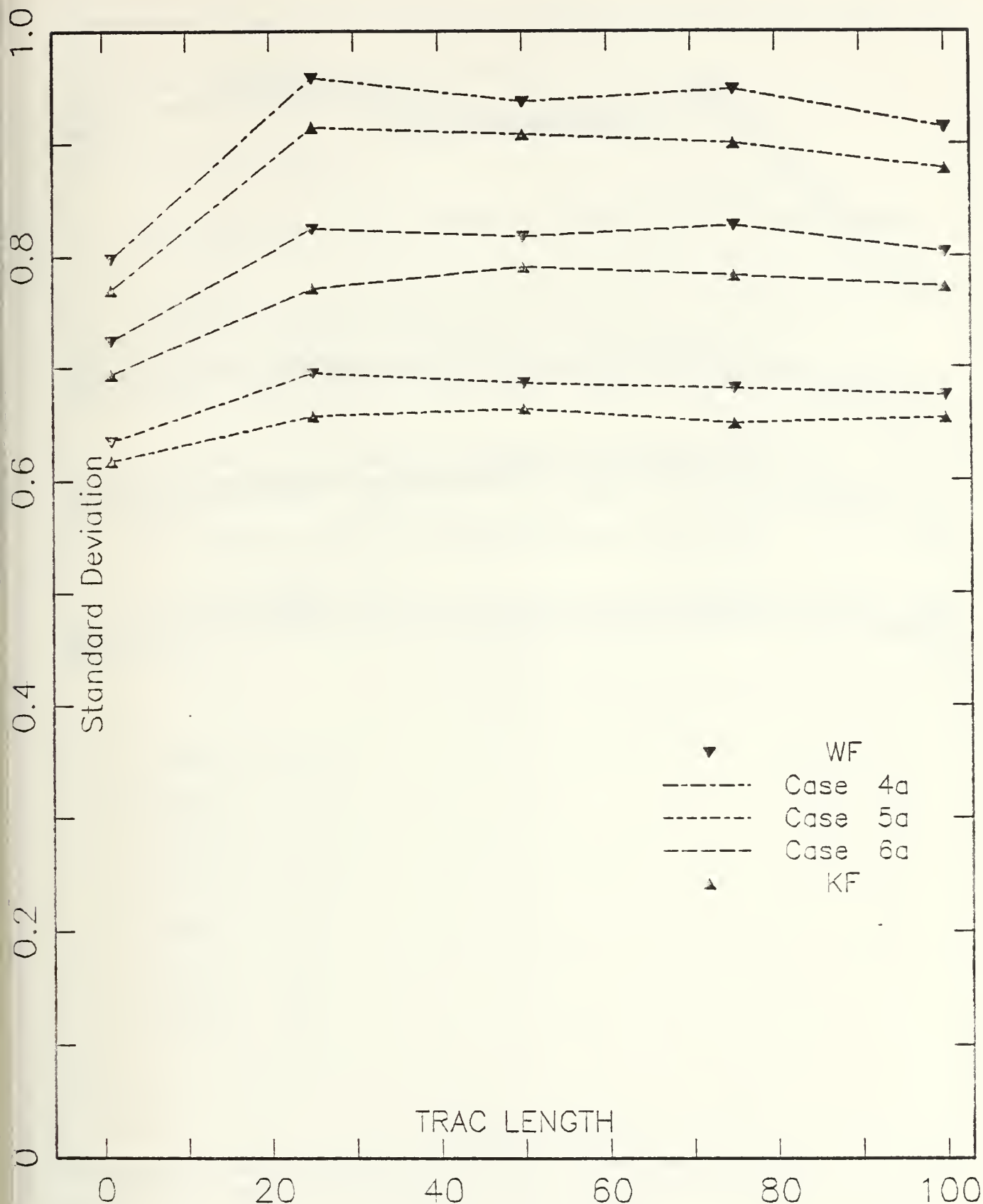
54

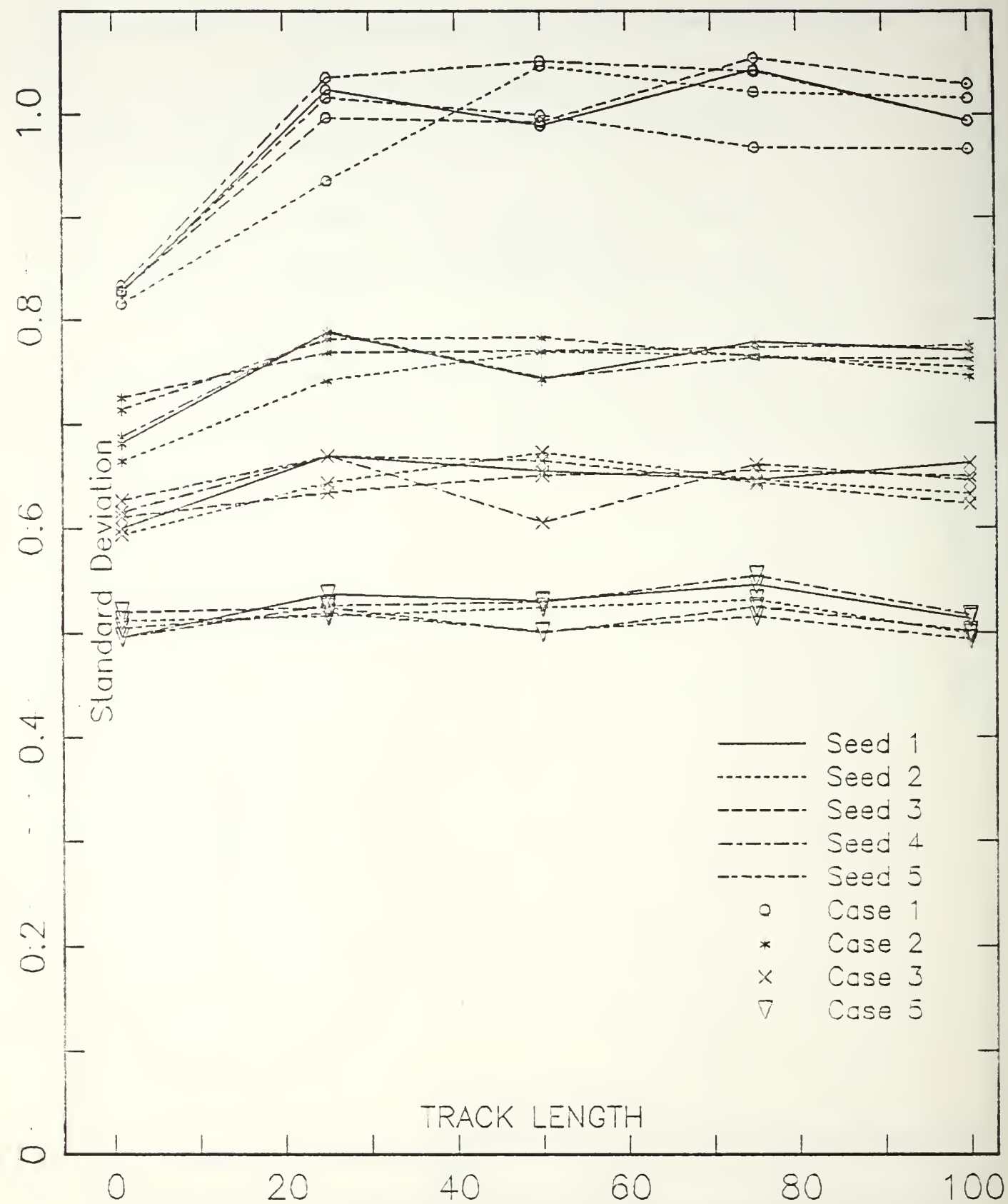Figure C.16   WF vs KF comparison for Case 4a-6a - Normal measurement errors.

Figure C.17   Variability of results.

56

LIST OF REFERENCES

1. Alan Washburn, *A Short Introduction to Kalman Filters*, Naval Postgraduate School. Handout

2. R.J.Meinhold and N.D.Singpurwalla, *Understanding the Kalman Filter*, The American Statistician:57. May 1983, Vol 37, No 2.

3. M.West, *Robust Sequential Approximate Bayesian Estimation*. J.R.Statist. Soc. B(1981), Vol 43, No 2.

4. D.P.Gaver and P.A.Jacobs, *Robustifying the Kalman Filter: Long Tailed Measurement Errors*, Unpublished Paper

5. B.J.T.Morgan, *Elements of Simulation*, Chapman and Hall 1986.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Professor D.P.Gaver Code 55 GV Naval Postgraduate School Monterey, CA 93943	3
4.	Professor P.A.Jacobs Code 55 JC Naval Postgraduate School Monterey, CA 93943	1
5.	Commander Amnon Shefi 1045 Eight Street Monterey, CA 93940	1
6.	Navy Attache Embassy of Israel, 3514 International Dr. N.W. Washington D.C. 20008	10
7.	LCDR Alexander Kukliansky 23/18 Levontin st. Netanya 42460 ISRAEL	5